Production of Polyketides by Saccharomyces cerevisiae

Wattanachaisaereekul, Songsak

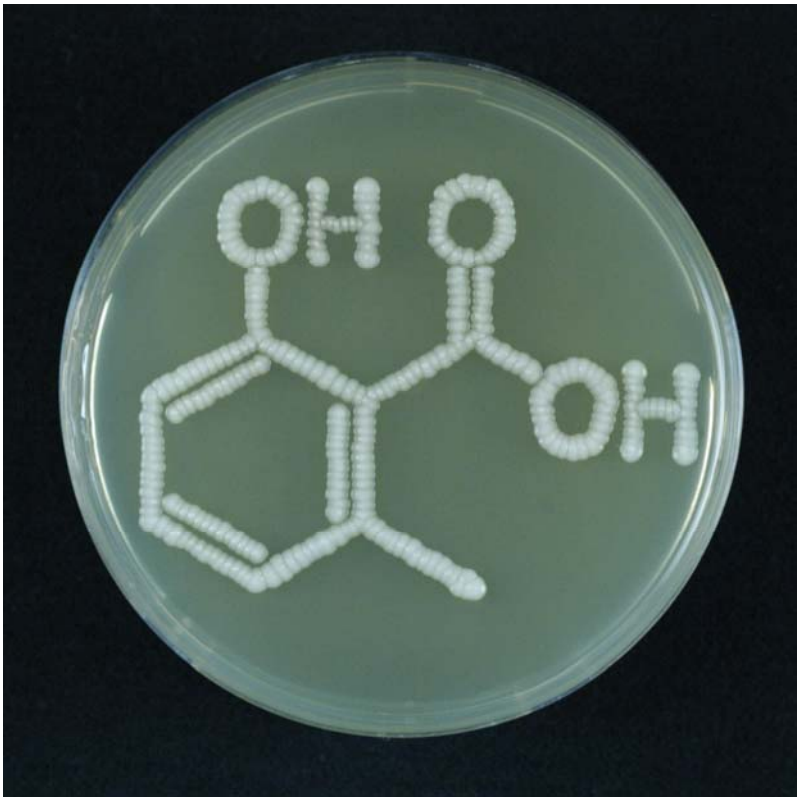
Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

Production of Polyketides by *Saccharomyces cerevisiae*

Songsak Wattanachaisaereekul

Ph.D. Thesis
September 2007



BioCentrum-DTU
TECHNICAL UNIVERSITY OF DENMARK

Production of Polyketides by *Saccharomyces cerevisiae*

Songsak Wattanachaisaereekul

Ph.D. Thesis

Center for Microbial Biotechnology
BioCentrum-DTU
Technical University of Denmark

Preface

This thesis represents the results of my Ph.D. study carried out at Center for Microbial Biotechnology, BioCentrum, Technical University of Denmark for the period of January 2003 to December 2006. This project was part of Euketides project (EU contract QLK3-CT-2002-01940).

Prof. Jens Nielsen and Associate Prof. Anna Eliasson Lantz have been my supervisors. I am very grateful for their excellent supervision, optimism and inspiration throughout the study, especially for Prof. Jens Nielsen who gave me the opportunity to carry out my Ph.D. study here. During the study, I have been working in the research area of molecular biology. I owe many thanks to Michael Lynge Nielsen for his supervision in that area. Part of this work was collaborated with the department of Biology, University of Iceland. Therefore, I would like to extend my gratitude to Prof. Ólafur S. Andrésson for his collaboration and valuable comments.

Co-operative help was rendered by Jette Mortensen and Tina Johansen while I was working with fermentations and HPLC. I also received valuable technical assistance and support from Lene Christiansen, Martin Nielsen, Peter Meincke, Elisabeth Krøger and Bettina Jørgensen. I thank you all for your help. The talented people, namely Prashant Bapat, Alessandro Fazio, Ana Paula Oliveira, Kiran Raosaheb Patil, and Jette Thykær are sincerely acknowledged for their help with experiments and data analysis. Special thanks to Kanchana Rueksomtawin, Wanwipa Vongsangnak and Pramote Chumnanpeun for their niceties and discussion. I also would like to thank all of my present and former colleagues at CMB for making this a wonderful place to work.

I wish to acknowledge the Ministry of Science and Technology, Thailand, for the Ph.D. grants throughout my study, and the Otto Mønsteds Fund for their financial support for my participation in a Natural Products Discovery and Production Conference in New Mexico, USA.

Last, but not least, I am profoundly grateful to my family and all of my dear friends for their support and encouragement. Without them, today would not have happened.

Songsak Wattanachaisaereekul

Summary

Polyketides (PKs) represent one of the largest groups of natural products found in fungi, bacteria and plants. Since many useful polyketides either originate from sources that are difficult or even impossible to cultivate or are produced in inadequate amounts in the natural hosts, it is of interest to express polyketide synthases (PKSs) in *Saccharomyces cerevisiae*. The advantages of this cell factory are that the techniques necessary for cultivation and genetic manipulation are well established and it is generally regarded as safe. 6-Methylsalicylic acid synthase (6-MSAS) was chosen as a model PKS which is naturally produced by e.g. *Penicillium patulum*. Functional expression of the 6-MSAS gene in *S. cerevisiae* requires the co-expression of a heterologous phosphopantetheinyl transferase (PPTase) to convert the apo-PKS into its active holo form.

In order to optimize the production of polyketides, several approaches were explored, e.g. using different PPTases from *B. subtilis* and *A. nidulans*, respectively, for the activation of heterologously expressed PKS in yeast. Metabolic engineering was also applied throughout this study e.g. by enhancing the supply of precursors for polyketide biosynthesis. Furthermore, the 6-MSAS gene was also integrated in the yeast chromosome to maintain a stable 6-MSA production, and different amounts of glucose were applied in the batch cultivations to investigate the optimum initial glucose concentration for the production of 6-MSA. The whole genome transcription was further analyzed together with the analysis of metabolic fluxes in order to reveal the differences between the reference strain and the strain expressing 6-MSAS. Finally, to identify further targets for metabolic engineering, an *in-silico* study using a genome scale metabolic model was applied for analysis of the effect on production of gene-deletions.

Dansk sammenfatning

Polyketider repræsenterer en stor gruppe af naturstoffer der producers af skimmelsvampe, bakterier og planter. En række af disses stoffer finder anvendelse som lægemidler, f.eks. antibiotika, kolesterolsænkende stoffer eller som kemoterapi i forbindelse med kræftbehandling. Mange interessante polyketider produceres dog naturligt i meget små mængder, og det er derfor interessant at udvikle nye generiske cellefabrikker til at producere disse stoffer i store mængder. Bagegær *Saccharomyces cerevisiae* er interessant at anvende som en sådan generisk cellefabrik idet der veludviklede teknikker til storskala fermentering af denne organisme, der er udviklet en række molekylærbiologiske værktøjer og den er såkaldt Generally Regarded as Safe (GRAS) godkendt. I dette studium er 6-Methylsalicylic acid (6-MSA) valgt som et model polyketid for at vurdere muligheden for at producere polyketider i gær. 6-MSA produceres normalt i skimmelsvampe, f.eks. *Penicillium patulum*, men for at kunne producere 6-MSA i gær kræver foruden ekspression af den polyketid synthetase der producerer 6-MSA også ekspression af en heterolog phosphopantetheinyl transferase (PPTase) der kan omdanne polyketid synthetasen fra dens inaktive apo-form til dens aktive holo-form.

For at optimere produktionen af 6-MSA i gær blev forskellige strategier evalueret. To forskellige PPTaser, fra henholdsvis *B. subtilis* og *A. nidulans*, blev evalueret, og det blev fundet af PPTasen fra *A. nidulans* gav højst produktivitet. Der blev endvidere foretaget metabolic engineering for at forøge dannelsen af precursoren malonoyl-CoA der anvendes til produktion af 6-MSA. Genet der koder for 6-MSA polyketid synthetasen blev også integreret i gærens kromosom for at etablere en stabil stamme. Til sidst i studiet blev nogle af de udviklede stammer analyseret i nærmere detaljer ved anvendelse af transkriptionsanalyse og metabolsk flux analyse.

Contents

Preface…… **i**

Summary……**iii**

Dansk sammenfatning…….…….. **v**

Contents……**vii**

Chapter 1 **Introduction**……**1**

1.1 Secondary metabolites……1

1.2 Polyketides……5

1.3 Biosynthesis of polyketides……... 8

1.4 6-Methylsalicylic acid, a model polyketide……..…....11

1.5 Heterologous expression of polyketide synthases……....16

1.6 Metabolic engineering of the production of polyketides in *S. cerevisiae*……..22

1.7 References……26

Chapter 2 **Optimization of heterologous production of the polyketide 6-MSA in *Saccharomyces cerevisiae*……37**

Abstract……...37

2.1 Introduction……...38

2.2 Materials and methods……40

2.2.1 Plasmids and strains……40

2.2.2 Molecular biology techniques……..41

2.2.3 Plasmid construction……41

2.2.4 Strain construction……... 42

2.2.5 Bioreactors and cultivation conditions……43
2.2.6 Absorbance and biomass measurement……43
2.2.7 Analysis of galactose, primary metabolites and 6-MSA……...44
2.3 Results……..44
2.3.1 Maximum specific growth rate, biomass yield and primary metabolites……...44
2.3.2 6-MSA production and yield……... 48
2.3.3 Bioinformatics of the PPTases……48
2.4 Discussion……49
2.5 References……52

Chapter 3 Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply……... 57
Abstract……..57
3.1 Introduction……..58
3.2 Materials and methods……59
3.2.1 Plasmids and strains……59
3.2.2 Bioreactors and cultivation conditions……61
3.2.3 Biomass measurement……. 62
3.2.4 Analysis of glucose, primary metabolites and 6-MSA……...62
3.3 Results……..63
3.3.1 Plasmid construction……63
3.3.2 Strain construction……... 65
3.3.3 Characterization of IBT100081, 100082 and 100083 in batch cultivations…… 66
3.3.4 Theoretical calculation of acetyl-CoA drain for 6-MSA production……... 71
3.3.5 The effect on 6-MSA production by a higher glucose concentration……...72

3.4 Discussion……73
3.5 References……75

Chapter 4 Towards stable production of 6-MSA by *S. cerevisiae* in chemostat cultivation……**79**
Abstract……79
4.1 Introduction……80
4.2 Materials and methods……82
4.2.1 Molecular biology techniques ……82
4.2.2 Plasmids and strains ……83
4.2.3 Construction of an integrating vector expressing 6-MSAS……84
4.2.4 Bioreactor and cultivation conditions……86
4.2.5 Absorbance and biomass measurement……87
4.2.6 Analysis of glucose, galactose and 6-MSA……87
4.3 Results……88
4.3.1 Characterization of the plasmid-based strain producing 6-MSA in chemostat cultivation ……88
4.3.2 Construction of a recombinant strain expressing 6-MSAS……91
4.3.3 Characterization of the 6-MSA integrated strain……92
4.4 Discussion……93
4.5 References……97

Chapter 5 Effect of glucose concentration on the production of the polyketide 6-MSA by *S. cerevisiae* in batch cultivations……**101**
Abstract……101
5.1 Introduction……102
5.2 Materials and methods……103
5.2.1 Strains……103
5.2.2 Bioreactors and cultivation conditions……103
5.2.3 Biomass measurement……104

5.2.4 Analysis of glucose, primary metabolites and 6-MSA……104
5.3 Results……105
5.3.1 Growth rate, production of 6-MSA and primary metabolites in the first exponential phase……..105
5.3.2 Growth rate, production of 6-MSA and primary metabolites in the second exponential phase……. 109
5.4 Discussion……..111
5.5 References……..114

Chapter 6 Transcriptional and metabolic responses to heterologous expression of a polyketide synthase in *S. cerevisiae*……117
Abstract……117
6.1 Introduction……118
6.2 Materials and methods……... 119
6.2.1 Strains…… 119
6.2.2 Bioreactors and cultivation conditions……...120
6.2.3 Sampling…… 121
6.2.4 Biomass measurement……... 122
6.2.5 Analysis of glucose, primary metabolites and 6-MSA……122
6.2.6 Extraction of total RNA……122
6.2.7 Probe preparation and hybridization……..123
6.2.8 Whole genome transcription data analysis…… 123
6.2.9 GC-MS analysis and fractional labeling of intracellular metabolites……124
6.2.10 Flux quantifications……... 125
6.3 Results and discussion……... 126
6.3.1 Physiological characterization.…….. 126
6.3.2 Transcription data and analysis……..127

6.3.3 Identification of reporter metabolites and metabolic subnetworks……...130
6.3.4 Global effect due to 6-MSA production…… 130
6.3.5 Global effect on amino acid metabolism……... 131
6.3.6 Global effect on fatty acid metabolism……..132
6.3.7 Summed fractional labeling……....137
6.3.8 Identification of pathways leading to pyruvate……..139
6.3.9 Metabolic flux calculation…… 141
6.4 References……..145

Chapter 7 *In silico* metabolic engineering for improved production of 6-MSA in *S. cerevisiae*……149
Abstract……149
7.1 Introduction……149
7.2 Materials and methods……... 150
7.2.1 Metabolic model…… 150
7.2.2 Flux balance analysis (FBA), minimization of metabolic adjustment (MOMA) and genetic algorithm……150
7.3 Results and discussion……... 150
7.4 References……..152

Chapter 8 Cloning, characterization and heterologous expression of *atX* from *A. terreus* IBT12713 and the hybrid 6-MSA Synthase gene from *P. patulum* and *A. terreus* in *S. cerevisiae*……...155
Abstract……155
8.1 Introduction……156
8.2 Materials and methods……... 158
8.2.1 Plasmids and strains……...158
8.2.2 Fungal cultivation and total RNA extraction…….158

8.2.3 Reverse transcription and polymerase chain reaction (RT-PCR) and yeast transformation…….159
8.2.4 Yeast cultivation and analysis of 6-MSA……...160
8.3 Results…….…...161
8.3.1 Construction of plasmid and strain…….161
8.3.2 Characterization of *atX* and heterologous production in *S. cerevisiae*…….161
8.3.3 Construction of the hybrid 6-MSAS gene from *P. patulum-A. terreus* and heterologous production in *S. cerevisiae*…….162
8.4 Discussion…….....163
8.5 References……..... 167

Conclusions and future perspectives……**169**

Appendix A Distribution of transcripts in Gene Ontology (GO) terms that are significantly changed in different pair-wise comparisons……..……...**173**
A.1 IBT100082 in ethanol vs glucose phase ……174
A.2 IBT100082 vs IBT100081 in the ethanol phase…… 176
A.3 IBT100082 vs IBT100081 in the glucose phase……177
A.4 IBT100083 vs IBT100082 in the ethanol phase…… 178

Appendix B The Sequence of *atX* Encoding 6-MSAS from *A. terrues* IBT12713 before and after Engineering……**179**
B.1 The sequence *atX* encoding 6-MSAS from *A. terreus* IBT12173 amplified by RT-PCR……..179
B.2 The engineered *atX* encoding 6-MSAS proposed by Java Codon Adaptation Tool……. 183

Appendix C Codon Usage in *Saccharomyces cerevisiae*……...**187**

Chapter 1

Introduction

1.1 Secondary metabolites

Microorganisms especially fungi and bacteria provide us with an enormous variety of secondary metabolites, several of which have profound biological activities that can be exploited. Secondary metabolites are often associated with differentiation and sporulation, and are therefore produced as soon as the growth rate of the producing micoorganism declines and during the stationary phase. This phase of growth has been termed the idiophase in contrast to the trophophase characterized by activity of the primary metabolism that is involved in synthesis of cellular macromolecules during exponential growth of microbes (Carlile et al., 2001). Hence, batch and fedbatch rather than continuous cultivation is usually favoured for secondary metabolite production.

Secondary metabolites tend to have an enormous variety of structures and biosynthetic origins, and they are often complex with unusual or unique components. There are several examples where the discovered secondary metabolites have been found to have structural features previously unknown in chemistry (Berdy, 1974). The tremendous variability in structures of secondary metabolites makes these compounds to belong to many different classes of organic compounds: aminocyclitols, amino sugars, quinones, coumarins, epoxides, glutarimides, glycosides, indole derivatives, lactones, macrolides, naphthalenes, nucleosides, peptides, phenazines, polyacetylenes, polyenes, pyrroles, quinolines, and terpenoids. In addition secondary metabolites possess unusual chemical linkages, such as β-lactam rings, cyclic peptides made of normal and modified amino acids, unsaturated bonds of polyacetylenes and polyenes, and large ring of macrolides. As a result of the broad substrate specificity of some biosynthetic enzymes, secondary metabolites are

produced typically as members of a particular chemical family (Martin and Gutierrez, 1992). In spite of the diversity in chemical structures, secondary metabolites have the property in common that they are all synthesized from low molecular weight compounds which also are intermediates in synthesis of cell constituents during exponential growth (Rose, 1976)

Although secondary metabolites do not have any obvious function in cell growth, it has been suggested that certain secondary metabolites may have a metabolic role, such as mediating passage of certain molecules across the plasma membrane (valinomycin, alamethicin), and playing a role in competition or combating other microbes in natural micro-environments (Rose, 1976), allowing the producer to invade territory occupied by the victim, and to utilize nutrients released upon its death (penicillin, cephalosporin, fusidic acid, griseofulvin, strobilurin, and sordarin) (Carlile et al., 2001). Additionally, in nature a secondary metabolites producer interacts with hundreds of other microorganisms and secondary metabolites are probably important signals in the communication between different members of the microbial communities (Davies, 2006). No molecular ecological studies are available to support or disclaim the notion of dispensability of secondary metabolites in nature. However, the variety of roles found for secondary metabolites suggests that removal of secondary metabolites from the producer organisms results in the long-term disappearance of those microorganisms from their usual habitats (Demain et al., 1989)

Secondary metabolites are greatly involved in the taxonomic studies and phylogenomic analysis, since the strains can be classified based on secondary metabolites profiles (Samson and Frisvad, 2004). However, the production of a particular secondary metabolite is restricted to certain taxonomic group of microorganisms which may not be closely related, and it varies considerably between different isolates of the same species. For example, cyclosporin A is produced by several strains of three species of *Tolypocladium*, but also by strains of species of *Beauveria*, *Fusarium*, *Neocosmospora* and *Stachybotrys* (Carlile et al., 2001).

The genes encoding the enzymes the in the biosynthesis pathways of secondary metabolite normally occur in clusters on a single chromosome. Examples includes ergot alkaloids in *Claviceps purpurea*, with 15 enzymic steps from tryptophan to ergotamine; aflatoxins in *Aspergillus parasiticus*; and penicillins in *Penicillium chrysogenum* (Bu'Lock and Barr, 1979; Sakuno et al., 2005). The significance of secondary metabolism gene clusters is that they may play a role in coordinated regulation of the expression of these genes, via appropriate transcription factors. In fungi, the genes in the same cluster are not under the same promoter but still likely to be co-regulated. In some cases a single polypeptide gene product may have multiple catalytic sites involved in the biosynthesis of a particular secondary metabolite (Carlile et al., 2001). The importance of secondary metabolites is easily defined by the numerous pharmaceutical agents that can be used for the treatment of infections, tumours and for promoting the health such as cholesterol lowering agents. Other important uses have also been found, for example versimide from *Aspergillus versicolor* which can be used as an insecticide (Cole and Rolinson, 1972), and monascorubrin from *Monascus* sp. that can be as a food colorant (Juzlova et al., 1996).

Secondary metabolites from micoorganisms in nature with different environments and climates have been screened continuously, e.g. the testing of culture broths from a large number of microbial isolates in specific biological or biochemical assays, ranging from the use of whole organisms to the use of single pure enzymes, for new bioactive compounds which may be developed to commercial products or are interesting to study for other purposes. Lots of new bioactive secondary metabolites were discovered from microorganisms living both pathogenically on animals and freely in the soils (Wawrik et al., 2007), glaciers (Zhao et al., 2007) and even in the deep ocean (Mayer et al., 2007). Many insect pathogenic fungi in the species of *Cordyceps* (Figure 1.1) and *Verticillium*, found in wild abundance tropical wet Thailand, represent a rich sources of structurally novel biologically active secondary metabolites that display the antimalarial and antituberculosis activities (Nilanonta et al., 2003; Isaka et al., 2005). More recently, a new bioactive compound exhibiting an antimicrobial property produced by *Roseobactor* sp. from the deep sea water was found during the Danish Galathea III expedition in 2006 (Figure 1.2).

Figure 1.1. The fungus *Cordyceps nipponica* grown on ant lions, collected from Khao Yai National Park, central Thailand, produces cordypyridones A and B which exhibit the antimalarial activity (Isaka et al., 2005).

Figure 1.2. *Roseobactor* sp., isolated from sea water produced the antibiotic polyketide indicated by the clear zone around the colonies. (Photograph: Rane Baadsgaard Lange, the Danish Expedition Foundation).

Microbial secondary metabolites represent an almost never ending source of bioactive compounds, with many compounds not yet being discovered. However, even though screening of microorganisms for the production of secondary metabolites has been very profitable, and has resulted in identification of many important pharmaceuticals, it is costly and time demanding. Thus, it takes typically between 8 and 15 years to develop a new drug or antibiotic. Only 1 in 15,000-50,000 from the original screening will make it to the marketplace. It costs more than 50 million dollars to bring a novel compound to market, involving initial research and development, testing for efficacy and safety, and clinical trials, but the rewards can be high. The secondary metabolite market is in excess of 40 billion dollars a year, as illustrated in Table 1.1.

Table 1.1. Estimated annual market of secondary metabolites (Demain, 2000)

Metabolite	World market ($ millions)
Antibiotics	28,000
Cephalosporins	11,000
Penicillins G & V	4,400
Erythromycins	3,500
Tetracyclines	1,400
Vancomycin & Teichoplanin	1,000
Rocephin (semisynthetic cephalosporin)	1,000
Hypocholesterolemic agents	
Statins	8,400
Immunosuppressants	
Cyclosporin A	1,500
Antihelmintics	
Avermectins	1,000
Antitumor agents	
Taxol	1,000
Bioinsecticides	
B_t toxin	125
Plant growth enhancers	
Gibberellins	120

1.2 Polyketides

Polyketides represent a major group of secondary metabolites with great structural variety (Figure 1.3), which derive from the condensation of acetate units resulting in the biosynthesis of a carbon chain, with alternate carbon atoms coming from the methyl and carboxyl groups of the acyl building block. The acetate origin of these compounds leads to a formation of even-numbered carbon chains (Mann, 1987; Hanson 2003). Several polyketide metabolites are of economic importance as antibiotics, anticancer drugs, immunosuppressants and cholesterol lowering agents such as erythromycin (macrolide antibiotic), vancomycin (aromatic antibiotic), rapamycin (polyene immunosuppressant), lovastatin (aromatic anticholesterol). Many polyketide metabolites such as aflatoxin, patulin, and ochratoxin A are known as mycotoxins causing diseases and contaminating stored food and animal feeds. Especially, the aflatoxins from *A. flavus* are among the most dangerous fungal toxins and cause liver cancer. (Miranda et al., 2007).

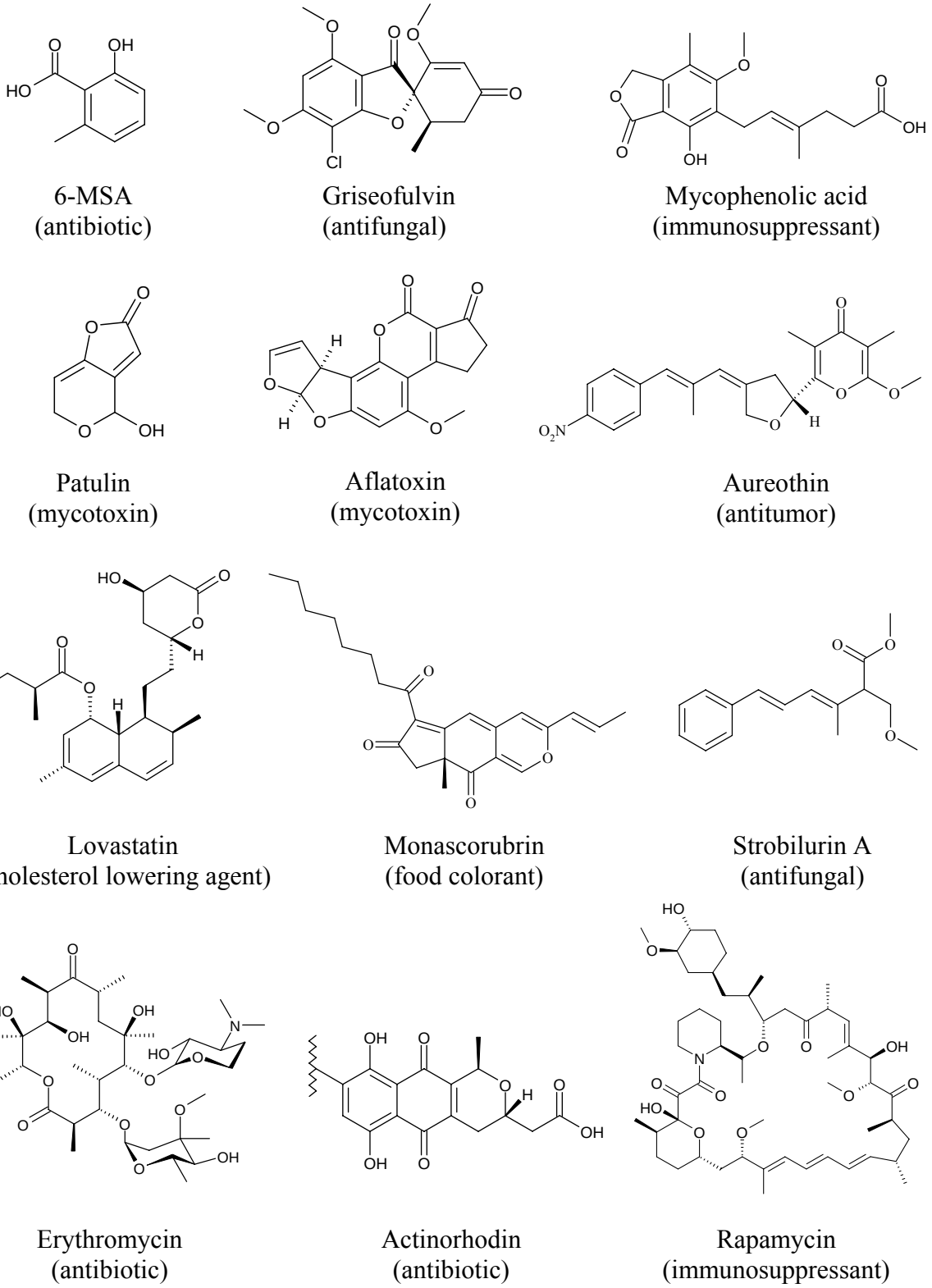


Figure 1.3. The structures of some selected polyketides.

The most significant group of prokaryotes which produce polyketides is the actinomycetes, whereas eukaryotic polyketide producers are primarily represented by not only the fungi, but also the plants producing chalcone metabolites which are believed to be involved in the defence against the invading fungal and bacterial plant pathogens (Martin, 1993). Polyketide biosynthesis is similar to the assembly of fatty acids which are considered as primary metabolites, since they are essential for the growth primarily as constituents of the cellular plasma membrane (Harwood, 1988; Ohlrogge et al, 1993; Smith, 1994). However, polyketides are classified as secondary metabolites, since these compounds, although they have cellular functions, are not essential to the survival of the organism. There are a great variety of polyketide secondary metabolites since they are synthesized by the diverse of prokaryotes and higher eukaryotes.

Even though the structural diversity of polyketides is tremendous, but they share a common origin, the decarboxylative condensation of malonate with an acylthioester to form a β-keto-thioester. This reaction is repeated depending on the programming of the enzymatic apparatus involved, leading to the condensation of acetate groups into a carbon chain containing ketogroups. The carbon backbones of all polyketide metabolites are assembled in this way, and it is the subsequent processing of the ketone group which is the key to the enormous structural diversity among the polyketides. If the ketone groups are reduced to methylene groups, this route leads to the formation of fatty acids. On the other hands, if there is no reduction of the keto groups the result will be the formation of aromatic compounds. However, many polyketides are results of processing of the carbon chain in between these two extremes, with some keto groups remaining unreduced, e.g. reduction only to a ketone or an alcohol group (Staunton and Weissman, 2001). One group of the polyketides are the aromatic ones and they may be further grouped by the number of their constituent acetate units ranging from triketide to more complex polyketides. Some typical aromatic microbial polyketides are the tetraketide 6-methylsalicylic acid (6-MSA) from *Penicillium patulum*, the immunosuppressive tetraketide mycophenolic acid from *P. brevicompactum*, the pentaketide mullein from *Aspergillus melleus*, the antifungal heptaketide griseofulvin from *P. griseofulvum*, and the nonaketides lovastatin from *A. terreus*, and aflatoxin from *A. flavus* (Turner, 1971).

1.3 Biosynthesis of polyketides

The enzymes responsible for polyketide production are called polyketide synthases (PKSs). Regardless of the differences in their end products, most PKSs share a minimum of four different domains, namely KS (ketosynthase), AT (acyltransferase), KR (ketoreductase), and ACP (acyl carrier protein). However, the domains DH (dehydratase), TE (thioesterase) and CYC (cyclase) are also found in PKSs. PKSs have been classified into three categories based on their enzyme architecture and gene organization, and how closely they mimic the architecture of type I (multifunctional polypeptide found in vertebrate) or type II (monofunctional and separated polypeptide with multienzyme complex found in bacteria and plant) fatty acid synthases. The three types of PKS are; type I) bacterial modular PKS e.g. erythromycin PKS, rapamycin PKS, and iterative fungal PKS e.g. 6-MSAS, lovastatin PKS, type II) bacterial aromatic PKS e.g. tetracycline PKS, actinorhodin PKS, and type III) plant PKS (Shen, 2003). The enzyme architectures of the first two types of PKSs are shown in Figure 1.4.

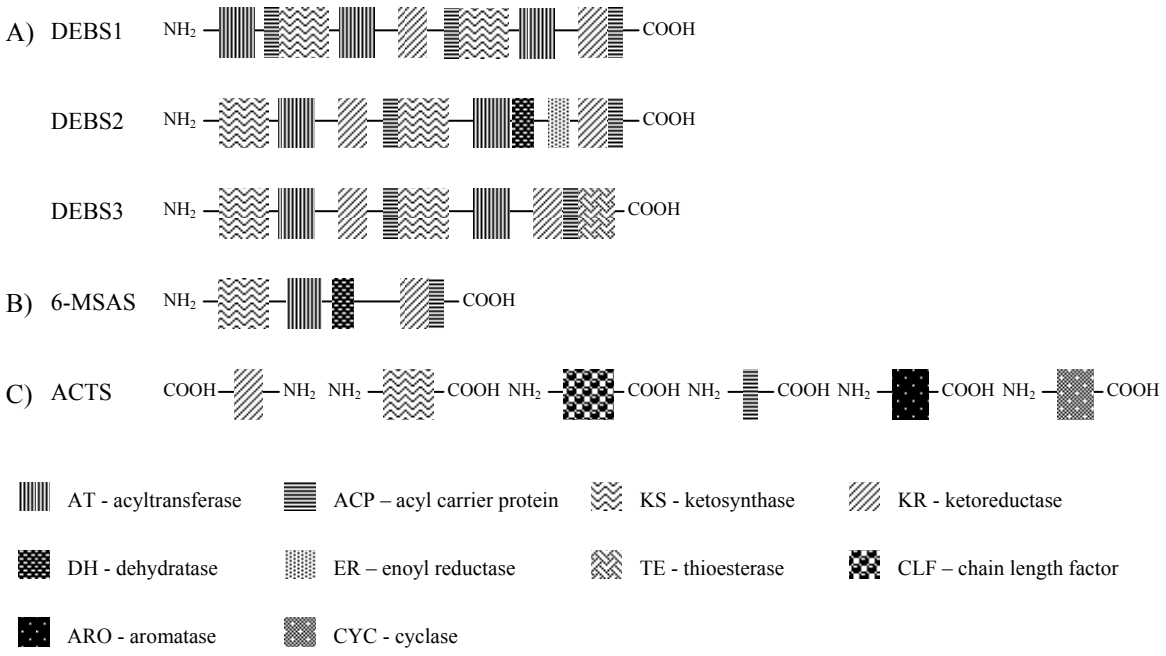


Figure 1.4. The domain architechture of polyketide synthases; A) type I bacterial modular PKS – erythromycin PKS; B) type I fungal PKS – 6-MSAS; C) type II bacterial aromatic PKS – actinorhodin PKS. (Figure modified from Staunton and Weissman, 2001)

Recently it was suggested to move away from the above mentioned classification of PKSs since there are some PKSs in the transitional stages of the different types (Müller, 2004). Regardless of type of PKS, there are highly conserved domains shared among all PKS genes, mostly notably within the KS domain (Bingle et al., 1999), and these domains may be useful for targeting PKSs in diverse organisms, as well as the studying the evolutionary and phylogenetic relationship between different PKSs.

The biosynthesis of polyketides is similar to that of fatty acids. The only difference is the degree of the subsequent reduction of the resulting keto groups as mentioned above. One of the keys to the control of polyketide or fatty acid biosynthesis is the ACP which introduces the intermediate acyl thioesters around the catalytic sites of the polyketide synthase, the acyl thioester being covalently attached to a 4´-phosphopantetheine moiety of the ACP which is post-translationally attached to a serine residue in the ACP (Hutchison et al. 1992). The biosynthesis of a polyketide is illustrated in Figure 1.5.

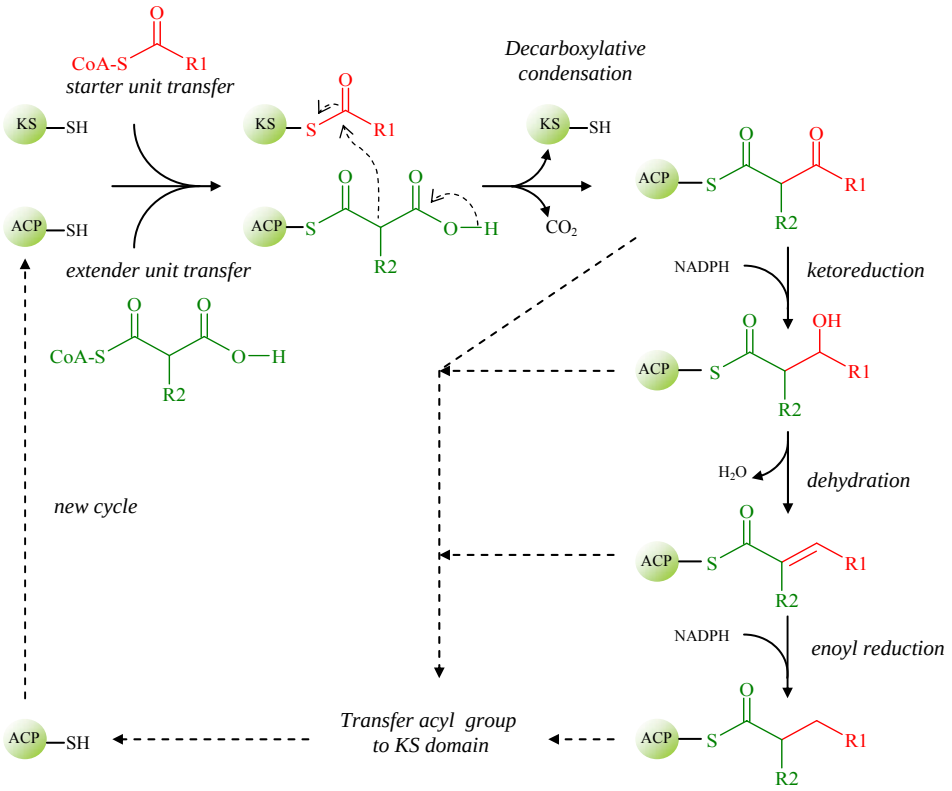


Figure 1.5. The biosynthesis of a polyketide. (Figure modified from Katz and Donadio, 1993)

The polyketide biosynthesis starts by the generation of malonyl-CoA from acetyl-CoA by acetyl-CoA carboxylase. Malonyl-CoA as an extender unit is then transacylated with the 4′-phosphopantetheine moiety of the ACP by acyl transferase (AT), resulting in the formation of malonyl-S-ACP. A molecule of acetyl-CoA as a starter unit is covalently attached to a cysteine residue at the active site of the condensing enzyme, ketosynthase (KS), followed by a decarboxylative condensation between the malonyl and the acetyl groups, catalyzed by the KS itself. The ketone of the resulting the ACP bound β-ketoacyl chain is then reduced, with one molecule of NADPH, by a keto reductase (KR) to a hydroxyl group (OH). The hydroxyl group could be subsequently dehydrated by dehydratase and further reduced by an enoyl reductase (ER) with the usage of one molecule of NADPH to yield the saturated butyryl intermediate. Finally the reduced butyryl moiety is transacylated to the cysteine residue on the KS domain and a new cycle can initiate – a malonyl group is attached to ACP and another cycle of condensation takes place, adding two carbon atoms to the growing acyl chain in each cycle. This process is repeated until the desired chain length is achieved. After that the chain will be released from the ACP by and acyl transferase or thioesterase. Whereas only malonyl-CoA is used as extender unit in aromatic polyketide, type I polyketides can use a variety of coenzyme A esters e.g. propionyl-CoA and methylmalonyl-CoA as building blocks. The use of different starter and extender units is allowing further diversity among polyketide structures in addition to the varying degree of reduction of the keto groups. After the synthesis of the polyketide chain, individual polyketides undergo subsequent modifications such as cyclizations, methylation, dimerizations or other highly specific modifications often called tailoring processes in order to reach the final products (Hopwood and Sherman, 1990).

1.4 6-Methylsalicylic acid, a model polyketide

Even though fungal polyketides represent a large group of secondary metabolites, only a few fungal polyketide-derived secondary metabolites have been studied in detail. These include mycotoxins from some species of *Aspergillus*, e.g. aflatoxin, patulin from *Penicillium* species (Steyn, 1981), some antimicrobial agents such as griseofulvin produced by *Penicillium griseofulvum*, polyketide-derived melanins from phytopathogenic fungi (Bell and Wheeler, 1986) and 6-methylsalicylic acid (6-MSA), a model polyketide, from *P. patulum*.

A significant advance in understanding fungal polyketide biosynthesis has been made through the study of 6-MSA biosynthesis by *P. patulum* (Shoolingin-Jordan and Campuzano, 1999). 6-MSA is one of the simplest polyketides since it is derived from one acetate starter molecule and three malonate extender units. (Martin and Gutierrez, 1992). The PKS responsible for biosynthesis of 6-MSA, so called 6-MSAS, contains the following domains; KS, AT, DH, KR and ACP (Figure 1.4). The biosynthesis of 6-MSA from *P. patulum* is shown in Figure 1.6, and it indicates that the domains in 6-MSAS are used repeatedly to catalyze three rounds of chain extension. The first condensation is followed directly by reaction with a second equivalent of malonate extender unit, while the second condensation is followed by reduction and dehydration of the newly-formed keto group. After the third cycle, the chain goes through cyclization, dehydration and enolisation to release the aromatic 6-MSA. The absence of a thioesterase (TE) domain in 6-MSAS could suggest the operation of a mechanism other than hydrolysis, involving a ketene intermediate, which could be achieved by the involvement of the adjacent hydroxyl group and hydration of the resulting ketene to the acid, although this has not been experimentally verified (Shoolingin-Jordan and Campuzano, 1999). 6-MSA, however, is later converted to patulin by the same fungus, *P. patulum*, resulting in low amounts of 6-MSA detected in the media, typically 0.2 mg/L in minimal media cultivations of *P. patulum* (Wattanachaisaereekul et al., unpublished data).

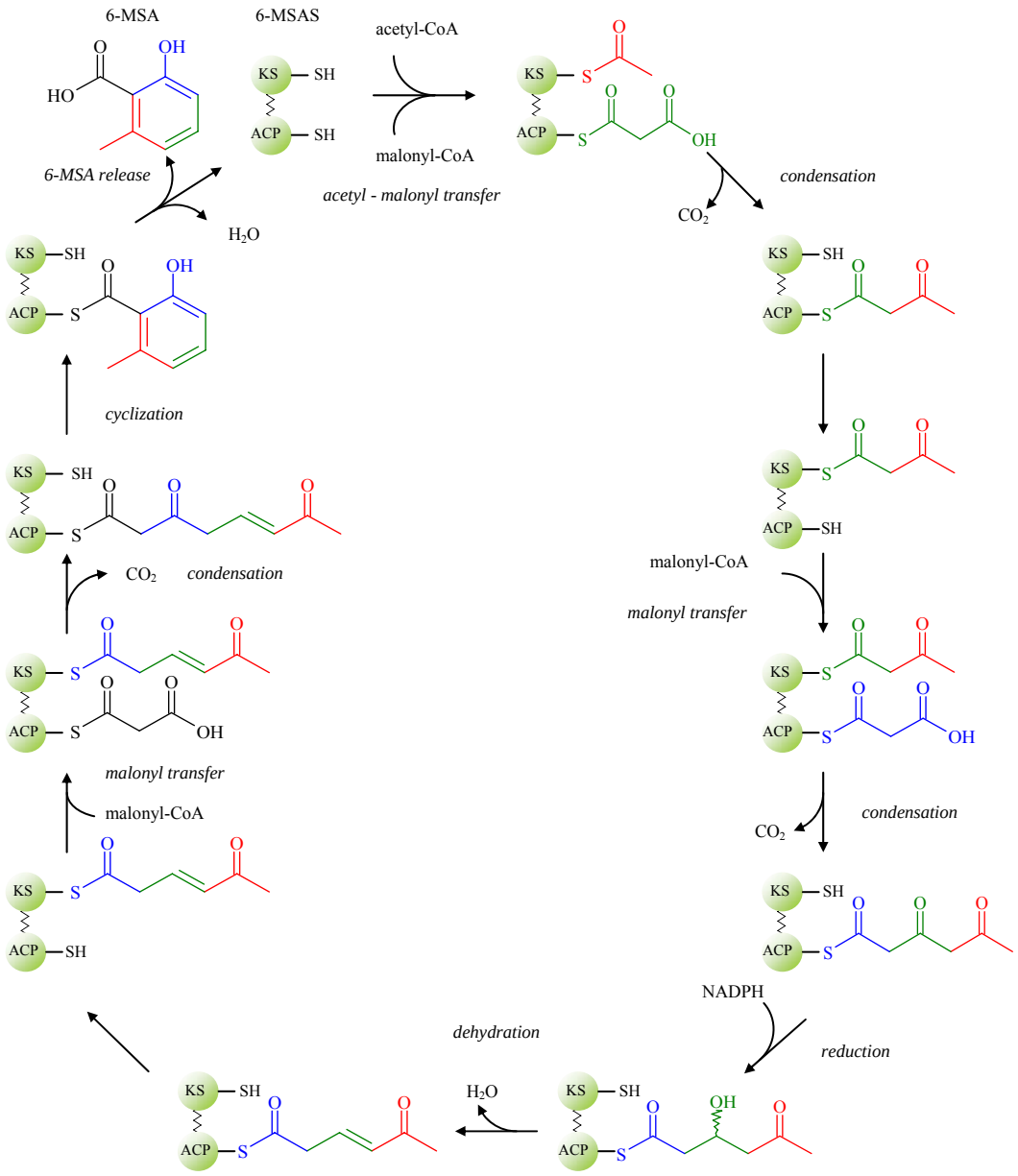


Figure 1.6. The biosynthetic pathway of the fungal polyketide 6-methylsalicylic acid (6-MSA) (Staunton and Weissman, 2001)

The 6-MSAS gene from *P. patulum* has been isolated and sequenced. It was identified as a 5,322 bp long open reading frame coding for a protein of 1,774 amino acids and 190,731 Da molecular mass (Beck et al., 1990). Additionally, one of the genes (*fas2*) encoding the fatty acid synthase in *P. patulum* is also know (Wiesner et al., 1988). Therefore, it is of interest to compare at the amino acid level between the fatty acid synthase (FAS) and the polyketide synthase (PKS) of the same fungus. Fatty acid synthase and other polyketide synthases occur either as large, multifunctional enzymes (type I PKS) or as aggregrated systems of individual component enzyme (type II PKS). In *Saccharomyces cerevisiae* and *P. patulum* the FAS is a multifunctional protein encoded by two unlinked gene FAS1 and FAS2 that encode β-subunit and α-subunit, respectively, which aggregate to form the α/β fatty acid synthase. (Lynen, 1980). The FAS2 proteins of *S. cerevisiae* and *P. patulum* are similar visualized by the domain organization in Figure 1.7 which shows that the acyl carrier protein (ACP), ketoreductase (KR) and ketosynthase (KS) functional regions are located on the large FAS2 polypeptide (1,887 amino acids). At the nucleotide level, the FAS2 gene of *P. patulum* reveals high similarity to the *S. cerevisiae* FAS2 genes (about 50-70% nucleotide similarity over the various domains) (Wiesner et al., 1988). However, the *P. patulum* FAS2 gene is interrupted by two short introns which are absent in the *S. cerevisiae* gene.

Inspection of the domain structure reveals that the organization of 6-MSAS gene from *P. patulum* is identical to that of *A. terreus* IBT12713 amplified by RT-PCR, which contained 5,409 bp coding for a protein of 1,803 amino acids (Chapter 7). Furthermore, comparison of the 6-MSAS amino acid sequence with type I FAS from *S. cerevisiae*, animals and its FAS2 shows that overall structure of 6-MSAS from *P. patulum* resembles more closely to that of animal FAS multifunctional proteins than the fungal α/β subunit type I FAS, indicating that the *P. patulum* 6-MSAS is more related and share a common evolutionary origin to a heterologous than to its own homologous FAS (Figure 1.7).

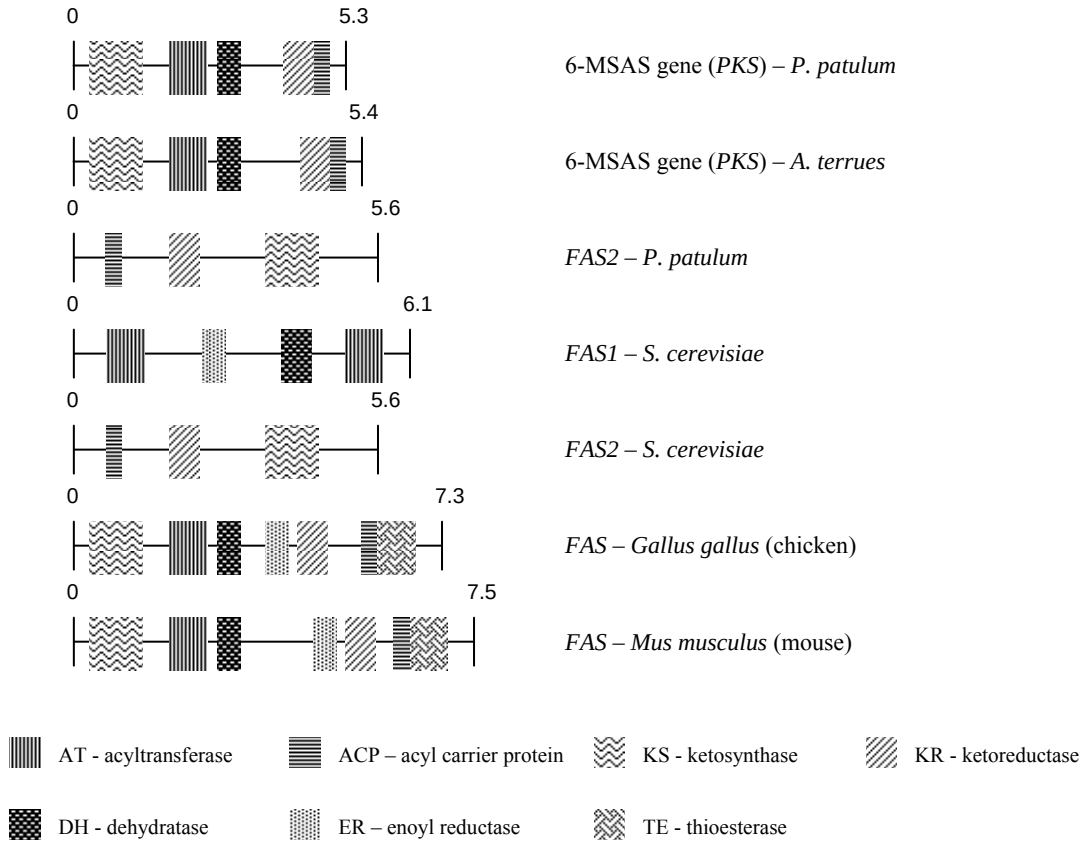


Figure 1.7. Organization of the domains in the 6-MSAS of *P. patulum* and *A. terrues* compared to the arrangement in the FAS of *P. patulum*, *S. cerevisiae*, chicken, and mouse from the conserved domain database (CDD). The number shown on the figure represents the length of the genes in kb.

From *P. patulum* 6-MSAS, a cysteine residue (Cysteine at position 204) was identified as the ketosynthase (KS) active site (Hopwood and Sherman, 1990) based on the similarity with the ketosynthase domains of various PKS, FAS, peptides and NRPS from different organisms. (Figure 1.8-A). The domain acyltransferase (AT) of the 6-MSAS (Serine at position 653) has been deduced by its similarity with the highly characteristic motif of the animal FAS acetyl-malonyl transferase domains. It seems as if the chain initiation and elongation of fungal polyketide is similar to that of the animal FAS system. However, the central region of the 6-MSAS has some similarity with the central region of the animal FAS and yeast FAS1, which encodes for the dehydratase domain.

In 6-MSAS of *P. patulum*, a ketoreductase (KR) domain exists near the C-terminus of the 6-MSAS (amino acid position 1419-1424) where a nucleotide biding site motif (Gly-Leu-Gyl-Val-Leu-Gly) fits the Gly-X-Gly-X-X-Gly consensus for the ketoreductase domains in the animal FAS. Finally, a predicted acyl carrier protein (ACP) region occurs at the C-terminus of the 6-MSAS. It includes a phosphopantetheine binding Serine at position 1732, which is similar to the phosphopantetheine biding sequence in several FASs, PKSs, and also in the oligopeptide synthetases from various animals and microorganisms (Figure 1.8-B).

A: KS domain

PKS			*
6-MSAS- *P.patulum*	195	GPSTAVDAACASSLVAIHHG	
6-MSAS- *A.terrues*	207	GPSTAVDAACASSLVAIHHG	
PKS- *B.subtilis*	671	GPSIAVSSACSSSLNAIHLA	
PKS- *Corynebacterium diphtheriae*	160	GPAFTVQTACSTSLVAIHLG	
PKS- Symbiont bacterium of Rove beetle	185	GPSLCVDTMCSGSLTALHLA	
PKS- *Streptomyces avermitilis*	201	GPSFAVDAACSSSLIALETA	
Melanin PKS- *Colletorichum lagenarium*	544	GPSLNVDTACSSSAAALNVA	
Red pigment bikaverin- *Gibberella fujikuroi*	191	GPSMTFDTACSASAVAIHTA	
CagA, Enediyne antitumor- *S.globisporus*	202	GGGYTVDGACSSSLLSITTA	
RedX- *S.coelicolor* A3(2)	717	AVPVAVEAACASSLAAVDVA	
Afb1, Aflatoxin B1- *A.parasiticus*	534	GPSYTNDTACSSSLAAIHLA	
MmpV, Mupirocin- *Pseudomonas fluorescens*	206	GPSLALDTMCSSAATALHLA	
YA, cornidial yellow pigment- *A.nidulans*	539	GPSVSVDTACSSSLAAIHLA	
FAS			
FAS- *Gallus gallus* (chicken)	152	GPSLTIDTACSSSLMALENA	
FAS- *Homo sapiens* (human)	152	GPSIALDTACSSSLMALQNA	
FAS- *Schizochytrium* sp. (marine micro-alga)	206	GMNCVVDAACASSLIAVKVA	
FASI- *Cryptosporidium parvum* (protozoan parasite)	3201	GPSMTVDTACASSLTAACIA	
FASNI- *Caenorhabditis elegans* (soil nematode)	161	GPSFSVDTACSSSLLALQLA	
PEPTIDE			
E.coli	158	GPCVTVQASCATSLVAVHLA	
Drosophila melanogaster (Fruit fly)	170	GPSFIVDTACSSSLVALNHA	
NRPS			
Albicidin pathotoxin synthetase- *X.albilineans*	897	GPSEVINSACSSSLVALHRA	
Pseudomonas syringae	1812	GPALSVQAACGSSLIAVHLA	
Yersinia pestis	832	GPALTIDTACSSSLVALDSA	

B: ACP domain

PKS			*
6-MSAS- *P.patulum*	1726	ADLGVDSVMTVTLRRQLQL	
6-MSAS- *A.terreus*	1756	SDLGVDSVMTVSLRGQLQK	
Curamycin PKS- *S.cyaneus*	38	ADFGLDSLGLLGIVGELEN	
OTC (Oxytetracyclin) PKS- *S.rimosus*	35	DALGYDSLALLNTVGRIER	
Monensin PKS- *S.cinnamonensis*	36	ALLGYESLALLETGGCIER	
PpsA (Phenolpthiocerol) PKS- *M. tuberculosis*	37	ADLGVSSRDAVVLSGELSE	
Tetracenomycin PKS- *S.glaucescens*	35	QDLGYDSIALLEISAKLEQ	
WA conidial green pigment- *A. nidulans*	1676	ADYGVDSLLSLTVTGKYRE	

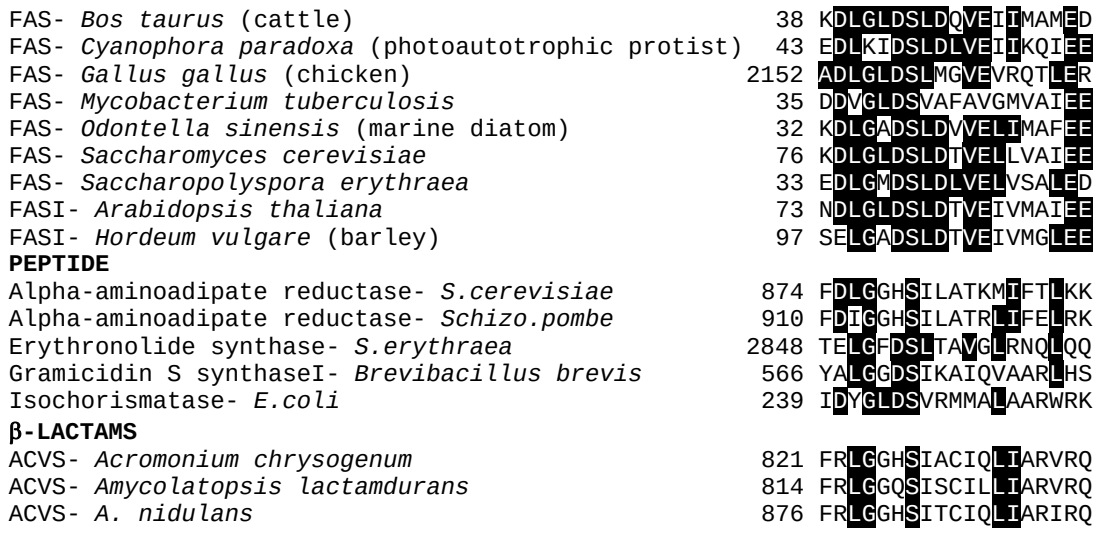


Figure 1.8. The conserved amino acid of 6-MSASs and several PKSs, FASs, and NRPSs in; A) KS domain, B) ACP domain. The conserved cystein (C) for starter unit binding site in KS domain, and conserved serine (S) for phosphopantetheine binding site in ACP domain are marked by *****. The sequences in this figure were obtained from the search of 6-MSAS gene in the conserved domain database (CDD).

1.5 Heterologous expression of polyketide synthases

The recombinant DNA technology that was introduced in the 1970s has made spectacular advances in the biotechnological and pharmaceutical industries. Many recombinant products are already on the market and many more are about to be commercialized. The knowledge of genetic and metabolic engineering technologies enables immense possibilities for manufacture of exotic and valuable substances that were virtually unobtainable before. These include the substances produced by non-native producers so-called heterologous products. Heterologous production is often associated with secondary metabolites such as various polyketides that are difficult to produce using the natural producers, and they are usually carried out in microorganisms or heterologous hosts mostly bacteria and yeast in order to achieve high yields by taking advantage of the rapid growth rates of these host cells.

The production of polyketides in heterologous hosts offers many advantages over the use of natural producers. As mentioned before, several useful polyketides are derived from the microorganisms with poor growth characteristics or even unculturable such as lichens mycobionts, which often grow a few millimeters per year in nature (Carlile et al., 2001) and approximately 800 secondary metabolites, of which more than half are polyketides, have been characterized from lichens (Huneck and Yoshimura, 1996). In order to identify and obtain these numerous hypothetical products and to use the enormous genetic potential in an efficient way, heterologous expression of these genes enconding PKS are required. In addition some of the natural producers yield poor titer of useful polyketides even in the optimized conditions. Therefore, the transfer of PKS gene from the unculturable sources or the sources with poor yields into a well-characterized and robust host is beneficial not only for enhancing the rate of product formation but also to make these compounds available for the pharmaceutical industry. Moreover, the use of more amenable hosts can even enhance the possibility to engineer novel polyketide analogs by combinatorial biosynthesis (Floss, 2006).

The examples of heterologous production of polyketides provided in Table 1.2 show that most of them are bacterial polyketide metabolites which are successfully produced in *Streptomyces* sp. and *E. coli*. For heterologous expression of bacterial PKSs gene, *Streptomyces* seems to be a better heterologous host than *E.coli*. Even though genetically modified PKS in *E. coli* can be more easily constructed and introduced, the uncommon precursor metabolites or unusual starter and extender units necessary for biosynthesis of the desired polyketide are often absent in *E. coli*.

Table 1.2. Examples of heterologous production of polyketides in different hosts.

Polyketide	Function	Natural producer	Heterologous host	Reference
Fungal type I PK				
6-MSA	antibiotic	*A. terreus*	*A. nidulans*	Fujii et al. (1996)
"	"	"	"	Regueira et al. (2007)
"	"	*P. patulum*	*S. cerevisiae, E.coli*	Kealey et al. (1998)
"	"	"	*S. cerevisiae*	Wattanachaisaereekul et al. (2007)*
"	"	"	*Nicotiana* spp. (Tobacco)	Yalpani et al. (2001)
Pentaketide T4HN	pigment	*C. lagenarium*	*A. oryzae*	Fujii et al. (1999)
Naphthopyrone	pigment	*A. fumigatus*	*A. oryzae*	Watanabe et al. (2000)
Lovastatin	anticholesterolemic	*A. terreus*	*E. coli*	Ma et al. (2006)
Bacterial type I PK				
6-deoxyerythronolide B	antibiotic	*Sacc. erythraea*	*E. coli*	Roberts et al. (1993)
"	"	"	"	Menzella et al. (2006)
"	"	"	*S. lividans*	Xue et al. (1999)
Rapamycin	Immunosuppressant	*S. hygroscopicus*	*S. lividans, E. coli*	König et al. (1997)
Picromycin/Methymycin	antibiotic	*S. venezuelae*	*S. lividans*	Tang et al. (1999)
Epothilone	antitumor	*Sor. Cellulosum*	*S. coelicolor*	Tang et al. (2000)
"	"	"	*E. coli*	Mutka et al. (2006)
Megalomicin	antiparasitic	*M. megalomicea*	*Sacc. erythraea*	Volchegursky et al. (2000)
Avermectin	anthelmintic	*S. avermitilis*	*S. lividans*	Wohlert et al. (2001)
Aureothin	antitumor, antifungal	*S. thioluteus*	*S. lividans*	He and Hertweck (2003), (2004)
Spinosyn	antibiotic	*Sacc. spinosa*	*Sacc. erythraea*	Martin et al. (2003)
Monensin	antibiotic	*S. cinnamonensis*	*S. coelicolor*	Oliynyk et al. (2003)
Soraphen A	antifungal	*Sor. cellulosum*	*S. lividans*	Zirkle et al. (2004)
Tylosin	antibiotic	*S. fradiae*	*S. venezuelae*	Jung et al. (2006)
Bacterial type II PK				
Enterocin	bacteriostatic	*S. maritimus*	*S. lividans*	Ziermann and Betlach (1999)
"	"	"	"	Hertweck et al. (2004)
Naphthocyclinone	antibiotic	*S. arenae*	*S. coelicolor*	Brünker et al. (1999), (2001)
Oleandomycin	antibiotic	*S. antibioticus*	*S. lividans*	Shah et al. (2000)
Elloramycin	antibiotic, antitumor	*S. olivaceus*	*S. lividans, E. coli*	Rafanan et al. (2001)
Maytansinoid	antitumor	*Actino. pretiosum*	*S. coelicolor*	Yu et al. (2002)
Actinorhodin	antibiotic	*S. coelicolor*	*S. lividans*	Kalaitzis and Moore (2004)
Fredericamycin	antitumor	*S. griseus*	*S. albus*	Wendt-Pienkowski et al. (2005)
Chartreusin	antitumor	*S. chartreusis*	*S. albus*	Xu et al. (2005)
Steffimycin	antitumor	*S. steffisburgensis*	*S. albus*	Gullón et al. (2006)
Oxytetracycline	antibiotic	*S. rimosus*	*S. coelicolor*	Zhang et al. (2006)
Plant type III PK				
Stibene	antifungal	*P. strobus* (pine)	*E. coli*	Raiber et al. (1995)
Naringenin	antioxidant	*R. idaeus* (raspberry)	*E. coli*	Zheng et al. (2001)
Anthocyanin	pigment	*Matthiola incana*	*E. coli*	Hemleben et al. (2004)

* See Chapter 2

Approximately 60% of the known bioactive polyketides are bacterial polyketides mainly from actinomycetes, and the rest are fungal polyketides. Due to the rudimentary knowledge about fungal polyketides, they have not been studied and heterologously produced in other hosts at the same level as bacterial polyketides, even though several polyketides from fungi exhibit interesting biological properties that cannot be found in bacteria. So far only a few fungal polyketides have been heterologously produced such as 6-MSA. Since this is a relatively small polyketide as well as it exhibits antibiotic property and is involved in the biosynthesis of the toxin patulin, 6-MSA was the first fungal polyketide to be cloned, purified and successfully produced in different kinds of hosts namely *E. coli*, *S. cerevisiae*, *S. coelicolor*, *A. nidulans*, and tobacco (Table 1.2). Much knowledge was gained from the biosynthesis and the heterologous production of 6-MSA, making 6-MSA a good choice as a model fungal polyketide.

Through the heterologous production of 6-MSA in each surrogate host, it was obvious that *E. coli* and *Streptomyces* are far from the ideal hosts for the production of fungal polyketides. It has been reported that eukaryotic genes or proteins expressed in prokaryotes are often incorrectly or incompletely processed after translation and wrongly folded, which results in formation of the wrong product (Primrose, 1986; Zabriskie and Arcuri, 1986). In *E. coli* the largest native protein or the RNA polymerase β-subunit is only 155 kDa in size (Julien et al., 2000) and it is therefore difficult to express large proteins like PKS in soluble and active form in *E. coli*. Furthermore, *E. coli* produces endotoxins and extensive purification is required to obtain products free of pyrogens (Kingsman et al., 1985). In addition, the use of bacteria as heterologous host for the production of fungal polyketide could be problematic due to the high GC-content in their chromosome especially in *Streptomyces* (Borodina et al., 2005), resulting in undesirable codon bias leading to low level translation.

For the heterologous production of fungal polyketides, a filamentous fungal expression system such as *A. nidulans* is better than bacterial systems. Since its metabolic function is well-characterized and there are good genetic tools for *A. nidulans*, this organism offers an alternative system for the expression of fungal polyketide synthases. Its native PPTase

(NpgA) is very efficient to phosphopantetheinylate PKSs at the ACP domain into the active form, and hence no PPTase from another microorgainsm is required. Furthermore, it is generally not necessary to remove the introns from PKS genes. However, the use of filamentous fungi as the heterologuous host for the production of fungal polyketide has so far been limited. This is due to the complicated biomolecular tools and techniques comparatively to that of other hosts. Furthermore, the wall growth and other undesirable morphologies that usually occur during cultivation of filamentous fungi in liquid media are problematic especially for large scale processes (Hamer, 2004).

Apart from the bacterial and filamentous fungi expression system, the budding yeast *S. cerevisiae* represents a useful alternative and is the most well characterized eukaryotic organism for expression of heterologous genes. A drawback on the use of *S. cerevisiae* as a heterologous host is that there is codon usage bias with expression of fungal genes. Furthermore, the PPTase (Ppt2 and Lys5) in *S. cerevisiae* which are involved only in fatty acid and lysine biosynthesis and not in the biosynthesis of polyketides (Stuible et al., 1998; Ehmann et al., 1999), cannot activate heterologously expressed PKSs. In other words, a PPTase has to be co-expressed in order to convert heterologous PKSs to their active form (Figure 1.9). In addition, the introns in fungal PKS encoding genes have to be removed as yeast can generally not remove fungal introns in its mRNA splicing system.

S. cerevisiae is a unicellular organism which is well suited for industrial fermentation and large scale processes. Unlike *E. coli* and *Streptomyces*, *S. cerevisiae* lacks detectable endotoxins and it is generally regarded as safe because of a long history of use in food and pharmaceutical industries. Genetic manipulation of *S. cerevisiae* has been made easier with greatly improved understanding of yeast biology and genetics. Recombinant genes can be expressed in this yeast correctly folded and modified at the posttranslational level to yield a biologically active product. Furthermore, the process technology for yeast cultures is well developed. Finally, expression in yeast may allow natural extracellular release of protein because of a secretion system that is similar to that of higher eukaryotes including native fungal polyketide producers. Biologically active secreted products are

substantially easier to recover than what is usually produced by recombinant bacteria (Chisti and Moo-Young, 1994).

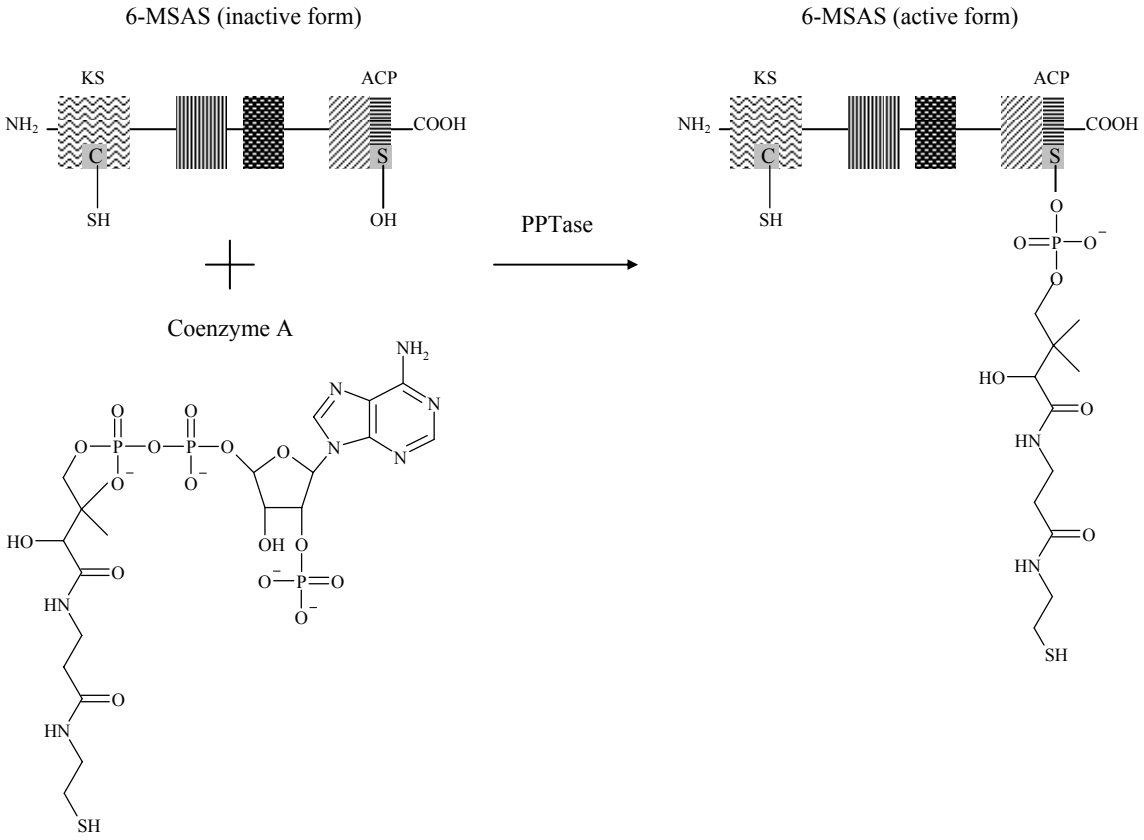


Figure 1.9. The change of 6-MSAS from an inactive into an active form by the transfer of phosphopantetheinyl moiety from Coenzyme A to the conserved serine (S) at ACP domain.

1.6 Metabolic engineering for the production of polyketide in *S. cerevisiae*

The metabolic engineering of polyketides has begun to prosper in the past decade due to the genomic research and the discovery of biosynthetic genes. While the biosynthetic pathways and genes for some polyketides have been known for many years, new polyketides have been found and known polyketides have been further optimized for higher production through metabolic engineering (Ryu et al., 2006).

Metabolic engineering is an approach which has developed to optimize the industrial fermentation processes and improve the cellular properties of the strains by introducing directed genetic changes in the cells, unlike the classical methods where strains are improved by e.g. random mutagenesis or genetic crossing (Nielsen, 2001). Metabolic engineering consists of analysis, design and synthesis (Figure 1.10). When the flux through the product is to be improved, the characterization or the analysis part of the currently applied strain such as fermentation physiology and numerous analytical techniques e.g. gene expression and flux analysis are performed. The data from the analysis part is interpreted and leads to the design for improving the properties of the strains. Subsequently, the synthesis part is performed according to the designed characteristics by using molecular biology techniques, gene cloning, transformation, gene deletion, and gene overexpression by either increasing the copy number or changing the strength of the promoter of the gene. The cycle is performed iteratively until the improved strain is achieved. According to Nielsen (2001), the aims of metabolic engineering can be categorized into the following groups:

1). Heterologous expression of genes in other organisms
2). Extension of substrate range and substrate utilization
3). Introduction of a new pathway or improvement of a pathway leading to a new product
4). Improvement of overall cellular physiology
5). Improvement of yield and productivity
6). Elimination or reduction of by-product formation

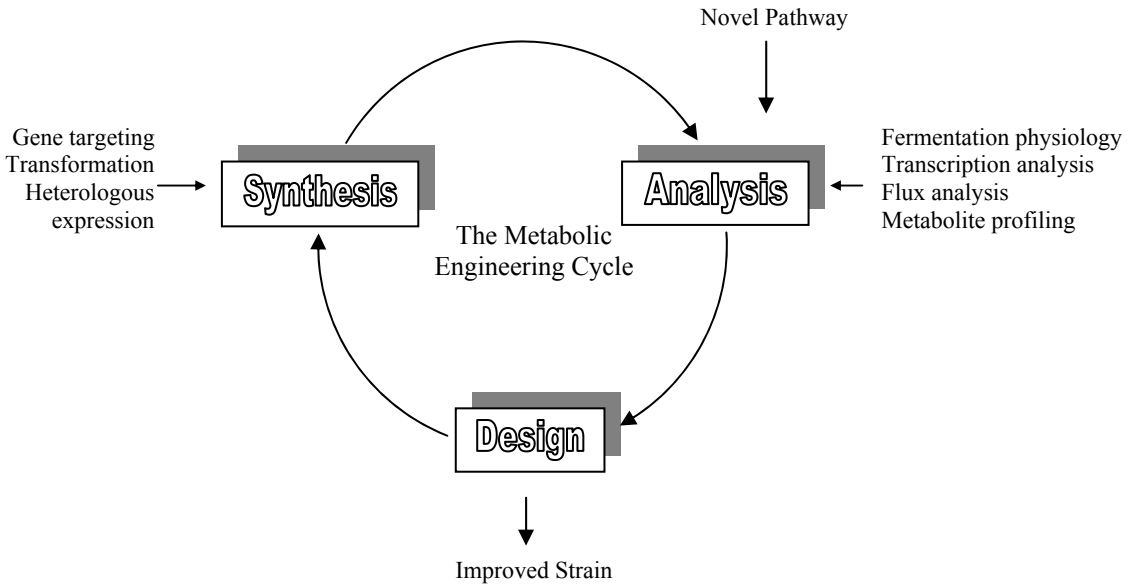


Figure 1.10. The metabolic engineering cycle (Nielsen, 2001).

There have been several examples of metabolic engineering of polyketide production which are mostly conducted in *Streptomyces* (McDaniel et al., 1994; Khosla, 1999; Zirkle et al., 2004; Floss, 2006). Only a few examples were performed in *S. cerevisiae* (Kealey et al., 1998; Mutka et al., 2006; Wattanachaisaereekul et al., 2007). This is likely due to the difficulty to express the PKS genes in yeast. However, except from the work presented in this thesis, metabolic engineering studies of polyketide production mostly dealt with the synthesis part and the important analysis part has largely been ignored, resulting in scarce knowledge about effects on strain characteristics and metabolism, hence lacking of information for the design of improved polyketide producers.

In order to evaluate the effect of the engineering and to analyze cellular metabolism, the metabolic regulation within the cell, which is complex and hierarchical (Figure 1.11), must be considered. Several high-throughput methods which cover large parts of an ‘‘ome’’ and are involved in the analytical part such as gas chromatography coupled with mass spectrometry for the analysis of metabolome and fluxome, 2D gel with mass spectrometry for proteome, and DNA microarray for transcriptome analysis have been developed and applied extensively in recent years in metabolic engineering studies of

yeast (Bro, 2003; Villas-Bôas, 2004). However, no reports related to analysis of heterologous polyketide production could be found in literature.

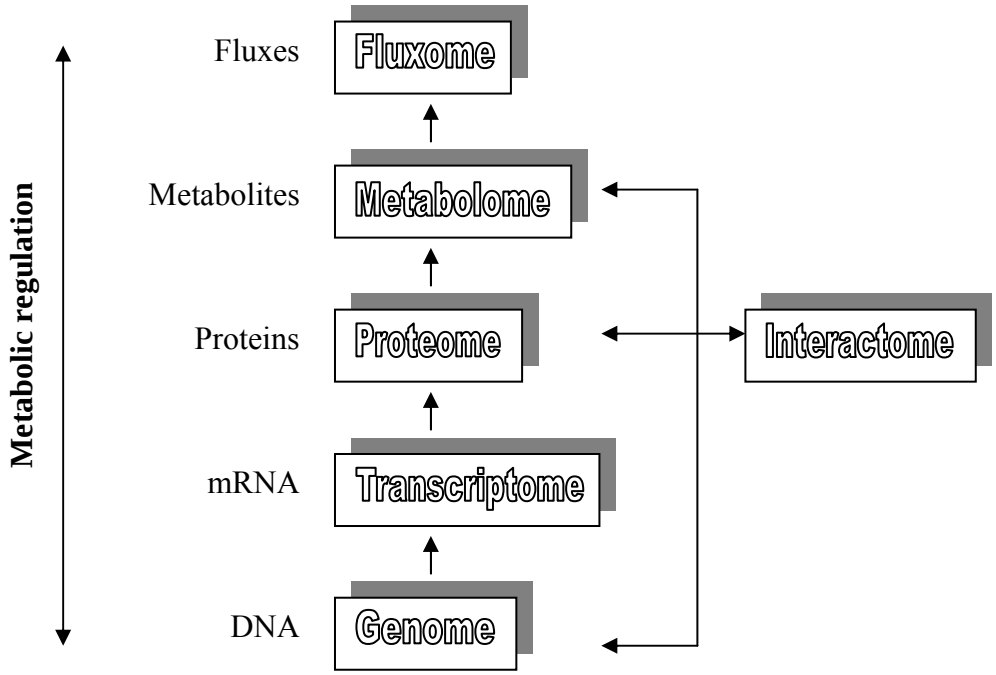


Figure 1.11. The scheme of the metabolic regulation in a cell. The genes in the genome are transcribed into mRNAs (transcriptome) which subsequently are translated into proteins (proteome). Proteins can interact with proteins themselves, DNA, or metabolites (interactome) and can be enzymes catalyzing the conversion to metabolites (metabolome). The conversion of metabolites results in the metabolic network, which can be quantified as the fluxes through metabolites (fluxome) (Bro, 2003).

A metabolic engineering approach was applied throughout the present Ph.D. study. The design and the synthesis part of metabolic engineering were involved in Chapter 2 and 3 where the heterologous expression of the 6-MSAS gene from *P. patulum* in *S. cerevisiae* was demonstrated. In Chapter 2, the 6-MSAS gene was under the control of *GAL1* promoter, and production of 6-MSA was optimized by using two different PPTase from *B. subtilis* and *A. nidulans*. The promoter of 6-MSA was then changed to an endogeneous glucose promoter in order to overcome the problem of slow growth on galactose and 6-

MSA could be produced during growth on glucose (Chapter 3). Moreover, the promoter of *ACC1* encoding acetyl-CoA carboxylase was replaced with the strong, constitutive *TEF1* promoter in order to ensure a high amount of malonyl-CoA that is used as a precursor for the biosynthesis of 6-MSA (Chapter 3). To improve the stable production of 6-MSA, the 6-MSAS gene was integrated in the yeast chromosome (Chapter 4). The production of 6-MSA was optimized in glucose minimal media with different initial glucose concentrations from 20-200 g/L in order to evaluate the optimum glucose concentration for the production of polyketide in yeast (Chapter 5).

Analysis of two ‘‘omes’’, transcriptome and fluxome of the cells were also performed in this study. The genome-wide transcription analysis by microarray technology was used to investigate the global transcriptional responses to introduction of the 6-MSAS gene and the recombinant strains were compared to the reference strain in batch cultivations (Chapter 6). Metabolic flux analysis was performed to further investigate the effect of the genetic modification on the fluxes through different parts of metabolism (Chapter 6). The high amount of information from the data analysis from Chapter 6 enables us to focus on systems of interacting biochemical reactions (a systems biology approach), instead of on only single pathways or reactions, which is definitely useful for designing a better strain and optimizing the production of polyketides in *S. cerevisiae*. Furthermore, an *in silico* study with a genome scale metabolic model of *S. cerevisiae* was applied in order to identify further targets for gene deletion for improving the production of 6-MSA (Chapter 7). Since the 6-MSAS gene from *P. patulum* was successfully expressed in yeast and high production was obtained, it was of interest to clone the 6-MSAS gene from another organism to investigate whether the production of 6-MSA is affected by the source of the gene. Therefore, the 6-MSAS gene from *A. terreus* was amplified and heterologous expressed in yeast. The outcome of this study is presented in Chapter 8.

1.7 References

Beck J, Ripka S, Siegner A, Schiltz E, Schweizer E. 1990. The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*: Its gene structure relative to that of other polyketide synthases. Eur. J. Biochem. 192: 487-498.

Bell AA, Wheeler, MH. 1986. Biosynthesis and functions of fungal melanins. Annual Review of Phytopathology 24: 411–451.

Berdy, J. 1974. Recent developments of antibiotic research and classification of antibiotics according to the chemical structure. Adv. Appl. Microbiol. 18: 309−406.

Bingle LEH, Simpson TJ, Larzarus CM. 1999. Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. Fungal Gen. Biol. 26: 209-223.

Borodina I, Krabben P, Nielsen J. 2005. Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. Genome Res. 15(6): 820-829.

Bro C. 2003. Transcription analysis of the yeast *Saccharomyces cerevisiae*. Ph.D. Thesis, BioCentrum-DTU, Technical University of Denmark.

Brünker P, McKinny K, Sterner O, Minas W, Bailey JE. 1999. Isolation and characterization of the naphthocyclinone gene cluster from *Streptomyces arenae* DSM 40737 and heterologous expression of the polyketide synthase genes. Gene 227: 125-135.

Brünker P, Sterner O, Bailey JE, Minas W. 2001. Heterologous expression of the naphthocyclinone hydroxylase gene from *Streptomyces arenae* for production of nevel hybrid polyketides. Antonie van Leeuwenhoek 79: 235-245.

Bu'Lock JD, Barr JG. 1979. The effects of maintained tryptophan levels on ergot alkaloid production and overall tryptophan economy in *Claviceps purpurea*. Biotechnol. Lett. 1(6): 261-265.

Carlile MJ, Watkinson SC, Gooday GW. 2001. The fungi (2nd edition). Academic Press. 588 p.

Chisti Y, Moo-Young M. 1994. Separation techniques in industrial bioprocessing. I. Chem. E. Symp. Ser. 137: 135-146.

Cole M, Rolinson GN. 1972. Microbial metabolites with insecticidal properties. Appl. Microbiol. 24(4): 660-662.

Davies J. 2006. Are antibiotics naturally antibiotics? J. Ind. Microbiol. Biotechnol. 33(7): 496-499.

Demain AL, 2000. Small bugs, big business: The economic power of the microbe. Biotechnology Advances 18: 499-514.

Demain AL, Somkuti GA, Hunter-Cevera JC, Rossmore HW. 1989. Novel microbial products for medicine and agriculture. Soc. Ind. Microbiol., Arlington. VA.

Ehmann DE, Gehring AM, Walsh CT. 1999. Lysine biosynthesis in *Saccharomyces cerevisiae*: mechanism of alpha-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. Biochemistry 38(19): 6171-6177.

Floss HG. 2006. Combinatiorial biosynthesis – Potential and problems. J. of Biotechnol. 124: 242-257.

Fujii I, Mori Y, Watanabe A, Kubo Y, Tsuji G, Ebizuka Y. 1999. Heterologous expression and product identification of *Colletotrichum lagenarium* polyketide synthase encoded by the *PKS1* gene involved in melanin biosynthesis. Biosci. Biotechnol. Biochem. 63(8): 1445-1452.

Fujii I, Ono Y, Tada H, Gomi K, Ebizuka Y, Sankawa U. 1996. Cloning of the polyketide synthase gene atX from *Aspergillus terreus* and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. Mol. Gen. Genet. 253: 1-10.

Gullón S, Olano C, Abdelfattah MS, Braña AF, Rohr J, Méndez C, Salas JA. 2006. Isolation, characterization, and heterologous expression of the biosynthesis gene cluster for the antitumor anthracycline steffimycin. Appl. Environ. Microbiol. 72(6): 4172-4183.

Hamer G. 2004. Wall growth in laboratory fermenters. Acta Biotechnologica 1(2): 139-143.

Hanson JR. 2003. Natural Products: the secondary metabolites. Tutorial chemistry texts 17th. The royal society of chemistry, UK. 147 p.

Harwood JL. 1988. Fatty acid metabolism. Annual review of plant physiology and plant molecular biology. 39: 101-138.

He J, Hertweck. 2003. Iteration as programmed even during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. Chem & Biol. 10: 1225-1232.

He J, Hertweck. 2004. Biosynthetic origin of the rare nitroaryl moiety of the polyketide antibiotic aureothin: Involvement of an unprecedented N-oxygenase. J. Am. Chem. Soc. 126: 3694-3695.

Hemleben V, Dressel A, Epping B, Lukačin, Martens S, Austin MB. 2004. Characterization and structural features of a chalcone synthase mutation in a white-flowering line of *Matthiola incana* R. Br. (Brassicaceae). Plant Mol. Biol. 55: 455-465.

Hertweck C, Xiang L, Kalaitzis JA, Cheng Q, Palzer M, Moore BS. 2004. Context-dependent behavior of the enterocin iterative polyketide synthase: A new model for ketoreductase. Chem & Biol. 11: 461-468.

Hopwood DA, Sherman DH. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu. Rev. Genet. 24: 37-66.

Huneck S, Yoshimura I. 1996. Identification of lichen substances (1st edition). Springer, Berlin. 493 p.

Hutchinson CR, Decker H, Guilfoile P, Shen B, Summers R, Wendt-Pienkowski E, Wessel B. 1992. Polyketide synthases: enzyme complexes and multifunctional proteins directing the biosynthesis of bacterial metabolites from fatty acids. *In* Petroski RJ, McCormick SP (eds.). Secondary metabolites: biosynthesis and metabolism. Plenum Press, New York. pp 3-10.

Isaka M, Kittakoop P, Kirtikara K, Hywel-Jones NL, Thebtaranonth Y. 2005. Bioactive substances from insect pathogenic fungi. Acc. Chem. Res. 38(10): 813-823.

Julien B, Shah S, Ziermann R, Goldman R, Katz L, Khosla C. 2000. Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*. Gene 249: 153-160.

Jung WS, Lee SK, Hong JSJ, Park SR, Jeong SJ, Han AR, Sohng JK, Kim BG, Choi CY, Sherman DH, Yoon YJ. 2006. Heterologous expression of tylosin polyketide synthase and production of a hybrid bioactive macrolide in *Streptomyces venezuelae*. Appl. Microbiol. Biotechnol. 72: 763-769.

Juzlova P, Martinkova L, Kren V. 1996. Secondary metabolites of the fungus Monascus: a review. J. Ind. Microbiol. 16(3): 163-170.

Kalaitzis JA, Moore BS. 2004. Heterologous biosynthesis of truncated hexaketides derived from the actinorhodin polyketide synthase. J. Nat. Prod. 67: 1419-1422.

Katz L, Donadio S. 1993. Polyketide synthesis: Prospects for hybrid antibiotics. Annu. Rev. Microbiol. 47: 875-912.

Kealey JT, Liu L, Santi DV, Betlach MC, Barr PJ. 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proc. Natl. Acad. Sci. 95: 505-509.

Khosla C. 1999. The generation of organic molecule diversity through metabolic engineering. *In* Lee SY, Papoutsakis ET (eds.). Metabolic engineering (2nd edition). Marcel Dekker, Inc. pp 203-225.

Kingsman SM, Kingsman AJ, Dobson MJ, Mellor J, Roberts NA. 1985. Heterologous gene expression in *Saccharomyces cerevisiae*. *In* Russell GE (ed.). Biotechnology and genetic engineering reviews, Intercept Ltd., Newcastle upon Tyne. pp 377-416.

König A, Schwecke T, Molnár I, Böhm GA, Lowden PAS, Staunton J, Leadlay PF. 1997. The pipecolate-incorporating enzyme for the biosynthesis of the immunosuppressant rapamycin: Nucleotide sequence analysis, disruption and heterologous expression of *rapP* from *Streptomyces hygroscopicus*. Eur. J. Biochem. 247: 526-534.

Lynen F. 1980. On the structure of fatty acid synthetase of yeast. Eur. J. Biochem. 112(3): 431-432.

Ma SM, Xie X, Wojcicki WA, Tang Y. 2006. Biosynthesis of lovastatin in a heterologous host. The abstract for Metabolic Engineering & Systems Biology Poster Session, The 2006 Annual meeting. 15 November, Grand Ballroom B, Hilton San Francisco, California, USA.

Mann J. 1987. Secondary metabolism. Oxford University Press, New York. 374 p.

Martin CJ, Timoney MC, Sheridan RM, Kendrew SG, Wilkinson B, Staunton J, Leadley PF. 2003. Heterologous expression in *Saccharopolyspora erythraea* of a pentaketide synthase derived from the spinosyn polyketide synthase. Org Biomol. Chem. 1: 4144-4147.

Martin CR. 1993. Structure, function, and regulation of the chalcone synthase. Int. Rev. Cyt. 147: 233-284.

Martin JF, Gutierrez. 1992. Molecular genetics of fungal secondary metabolites. *In* Kinghorn JR, Turner G (eds.). Applied molecular genetics of filamentous fungi. Blackie Academic & Professional, Glasgow. 259 p.

Mayer AM, Rodríguez AD, Berlinck RG, Hamann MT. 2007. Marine pharmacology in 2003-4: marine compounds with anthelmintic antibacterial, anticoagulant, antifungal, anti-inflammatory, antimalarial, antiplatelet, antiprotozoal, antituberculosis, and antiviral

activities; affecting the cardiovascular, immune and nervous systems, and other miscellaneous mechanisms of action. Comp Biochem Physiol C Toxicol Pharmacol. 145(4): 553-81.

McDaniel R, Ebert-Khosla S, Fu H, Hopwood DA, Khosla C. 1994. Engineered biosynthesis of novel polyketides: Influence of a downstream enzyme on the catalytic specificity of a minimal aromatic polyketide synthase. Proc. Natl. Acad. Sci. 91: 11542-11546.

Menzella HG, Reisinger SJ, Welch M, Kealey JT, Kennedy J, Reid R, Tran CQ, Santi DV. 2006. Redesign, synthesis and functional expression of the 6-deoxyerythronolide B polyketide synthase gene cluster. J. Ind. Microbiol. Biotechnol. 33: 26-28.

Miranda DDC, Arçari DP, Ladeira MSP, Calori-Domingues MA, Romero AC, Salvadori DMF, Gloria EM, Pedrazzoli J, Ribeiro ML. 2007. Analysis of DNA damage induced by aflatoxin B1 in Dunkin-Hartley guinea pigs. Mycopathologia. 163(5): 275-280.

Mutka SC, Carney JR, Liu Y, Kennedy J. 2006. Heterologous production of epothilone C and D in *Escherichia coli*. Biochem. 45: 1321-1330.

Müller R. 2004. Don't classify polyketide synthases. Chem & Biol. 11: 4-6.

Nielsen J. 2001. Metabolic engineering. Appl. Microbiol. Biotechnol. 55: 263-283.

Nilanonta C, Isaka M, Chanphen R, Thong-orn N, Tanticharoen M, Thebtaranonth Y. 2003. Unusual enniatins produced by the insect pathogenic fungus *Verticillium hemipterigenum*: isolation and studies on precursor-directed biosynthesis. Tetrahedron. 59: 1015-1020.

Ohlrogge JB, Jaworski JG, Post-Beittenmiller D. 1993. De novo fatty acid biosynthesis. *In* Morre TS (ed.). Lipid Metabolism in Plants. CRC press, Boca Raton. pp 3-32.

Oliynyk M, Stark CBW, Bhatt A, Jones MA, Hughes-Thomas ZA, Wilkinson C, Oliynyk Z, Demydchuk Y, Staunton J, Leadlay PF. 2003. Analysis of the biosynthetic gene cluster

for the polyether antibiotic monensin in *Streptomyces cinnamonensis* and evidence for the role of *monB* and *monC* genes in oxidative cyclization. Mol. Microbiol. 49(5): 1179-1190.

Primrose SB. 1986. The application of genetically engineered micro-organisms in the production of drugs. J. Appl. Bacteriol. 61: 99-116.

Rafanan ER, Le L, Zhao L, Decker H, Shen B. 2001. Cloning, sequencing, and heterologous expression of the *elmGHIJ* genes involved in the biosynthesis of the polyketide antibiotic elloramycin from *Streptomyces olivaceus* Tü2353. J. Nat. Prod. 64: 444-449.

Raiber S, Schröder G, Schröder J. 1995. Molecular and enzymatic characterization of two stilbene synthases from eastern white pine (*Pinus strobus*): a single Arg/His difference determines the activity and the pH dependence of the enzymes. FEBS Lett. 361: 299-302.

Regueira TB. 2007. Discovery, cloning and heterologous expression of novel fungal polyketide synthases. Ph.D. Thesis, BioCentrum-DTU, Technical University of Denmark.

Roberts GA, Staunton J, Leadlay PF. 1993. Heterologous expression in Escherichia coli of an intact multienzyme component of the erythromycin-producing polyketide synthase. Eur. J. Biochem. 214: 305-311.

Rose AH. 1976. Chemical Microbilogy: An introduction to Microbial Physiology (3rd edition). Plenum Press, New York. 469 p.

Ryu YG, Butler MJ, Chater KF, Lee KJ. 2006. Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*. Appl Environ Microbiol. 72(11): 7132-7139.

Sakuno E, Wen Y, Hatabayashi H, Arai H, Aoki C, Yabe K, Nakajima H. 2005. *Aspergillus parasiticus* cyclase catalyzes two dehydration steps in aflatoxin biosynthesis. Appl. Environ. Microbiol. 71(6): 2999-3006.

Samson RA, Frisvad JC. 2004. *Penicillium* subgenus *Penicillium*: new taxonomic schemes and mycotoxins and other extrolites. Studies in Mycology 49. An institute of the Royal Netherlands Academy of Arts and Sciences.

Shah S, Xue Q, Tang L, Carney JR, Betlach M, MaDaniel R. 2000. Cloning, characterization and heterologous expression of a polyketide synthase and P-450 oxidase involved in the biosynthesis of the antibiotic oleandomycin. J. Antibiot. 53(5): 502-508.

Shen B. 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Current Opinion in Chemical Biology. 7(2): 285-295.

Shoolingin-Jordan PM, Campuzano IDG. 1999. Biosynthesis of 6-Methysalicylic Acid. *In* Barton D, Nakanishi K (eds-in-chief), Kelly JW (ed.). Comprehensive Natural Products Chemistry. 1(14): 345-365

Smith S. 1994. The animal fatty acid synthase: One gene, one polypeptide, seven enzyme. FASEB J. 8: 1248-1259.

Staunton J, Weissman KJ. 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18: 380-416.

Steyn PS. 1981. The biosynthesis of mycotoxins: A study in secondary metabolism. Mycologia. 73(4): 792-793.

Stuible HP, Meier S, Wagner C, Hannappel E, Schweizer E. 1998. A novel phosphopantetheine: protein transferase activating yeast mitochondrial acyl carrier protein. J. Biol. Chem. 273(35): 22334-22339.

Tang L, Fu H, Betlach MC, McDaniel R. 1999. Elucidating the mechanism of chain termination switching in the picromycin/mythymycin polyketide synthase. 1999. Chem & Biol. 6(8): 553-558.

Tang L, Shah S, Chung L, Carney J, Katz L, Khosla C, Julien B. 2000. Cloning and heterologous expression of the epothilone gene cluster. Science 287: 640-642.

Turner WB. 1971. Fungal metabolites. Academic Press, London. 446 p.

Villas-Bôas SG. 2004. Metabolome analysis of the yeast *Saccharomyce cerevisiae*. Ph.D. Thesis. BioCentrum-DTU, Technical University of Denmark.

Volchegursky Y, Hu Z, Katz L, McDaniel R. 2000. Biosynthesis of the anti-parasitic agent megalomicin: transformation of erythromycin to megalomicin in *Saccharopolyspora erythraea*. Mol. Microbiol. 37(4): 752-762.

Watanabe A, Fujii I, Tsai HF, Chang YC, Kwon-Chung KJ, Ebizuka Y. 2000. *Aspergilllus fumigatus alb1* encodes naphthopyrone synthase when expressed in *Aspergillus oryzae*. FEMS Microbiol. Lett. 192: 39-44.

Wattanachaisaereekul S, Lantz AE, Nielsen ML, Andrésson ÓS, Nielsen J. 2007. Optimization of heterologous Production of the polyketide 6-MSA in *Saccharomyces cerevisiae*. Biotechnol Bioeng. 97(4): 893-900.

Wawrik B, Kutliev D, Abdivasievna UA, Kukor JJ, Zylstra GJ, Kerkhof L. 2007. Biogeography of Actinomycete Communities and Type II Polyketide Synthase Genes in Soils Collected in New Jersey and Central Asia. App. Environ. Microbiol. 73(9): 2982-2989.

Wendt-Pienkowski E, Huang Y, zhang J, Li B, Jiang H, Kwon H, Hutchinson CR, Shen B. 2005. Cloning, sequencing, analysis, and heterologous expression of the fredericamycin biosynthetic gene cluster from *Streptomyces griseus*. J. Am. Chem. Soc. 127: 16442-16452.

Wiesner P, Beck J, Beck KF, Ripka S, Muller G, Lucke S, Schweizer E. 1988. Isolation and sequence analysis of the fatty acid synthetase FAS2 gene from *Penicillium patulum*. Eur. J. Biochem. 177(1): 69-79.

Wohlert SE, Lomovskaya N, Kulowski K, Fonstein L, Occi JL, Gewain KM, MacNiel DJ, Hutchinson CR. 2001. Insights about the biosynthesis of the avermectin deoxysugar L-

oleandrose through heterologous expression of *Streptomyces avermitilis* deoxysugar genes in *Streptomyces lividans*. Chem & Biol. 8: 681-700.

Xu Z, Jakobi K, Welzel, Hertweck. 2005. Biosynthesis of the antitumor agent chartreusin involves the oxidative rearrangement of an antracyclic polyketide. Chem & Biol. 12: 579-588.

Xue Q, Ashley G, Hutchinson CR, Santi DV. 1999. A multiplasmid approach to preparing large libraries of polyketides. PNAS. 96(21): 11740-11745.

Yalpani N, Altier DJ, Barbour E, Cigan AL, Scelonge CJ. 2001. Production of 6-methylsalicylic acid by expression of a fungal polyketide synthase activates disease resistance in tobacco. The Plant Cell. 13: 1401-1409.

Yu TW, Bai L, Clade D, Hoffmann D, Toelzer S, Trinh KQ, Xu J, Moss SJ, Leistner E, Floss HG. 2002. The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*. PNAS. 99(12): 7968-7973.

Zabriskie DW, Arcuri EJ. 1986. Factors influencing productivity of fermentations employing recombinant microorganisms. Enzyme Microb. Technol. 8: 706-717.

Zhang W, Watanabe K, Wang CCC, Tang Y. 2006. Heterologous biosynthesis of amidated polyketides with novel cyclization regioselectivity from oxytetracycline polyketide synthase. J. Nat. Prod. 69: 1633-1636.

Zhao J, Yang N, Zeng R. 2007. Phylogenetic analysis of type I polyketide synthase and nonribosomal peptide synthetase genes in Antarctic sediment. Extremophiles. DOI 10.1007/s00792-007-0107-9.

Zheng D, Schröder G, Schröder J, Hrazdina G. 2001. Molecular and biochemical characterization of three aromatic polyketide synthase genes from *Rubus idaeus*. Plant Mol. Biol. 46: 1-15.

Ziermann R, Betlach MC. 1999. Recombinant polyketide synthesis in Streptomyces: Engineering of improved host strains. Biotechniques 26: 106-110.

Zirkle R, Ligon JM, Molnár I. 2004. Heterologous production of the antifungal polyketide antibiotic soraphen A of *Sorangium cellulosum* So ce26 in *Streptomyces lividans*. Microbiol. 150: 2761-2774.

Chapter 2

Optimization of Heterologous Production of the Polyketide 6-MSA in *Saccharomyces cerevisiae*

Songsak Wattanachaisaereekul[1], Anna Eliasson Lantz[1*], Michael Lynge Nielsen[1],
Ólafur S. Andrésson[2], and Jens Nielsen[1]

Biotechnology and Bioengineering 97(4): 893-900.

Abstract

Polyketides are a group of natural products that have gained much interest due to their use as antibiotics, cholesterol lowering agents, immunosuppressors and as other drugs. Many organisms that naturally produce polyketides are difficult to cultivate and only produce these metabolites in small amounts. It is therefore of general interest to transfer polyketide synthase (PKS) genes from their natural sources into heterologous hosts that can over-produce the corresponding polyketides. In this study we demonstrate the heterologous expression of 6-methylsalicylic acid synthase (6-MSAS), naturally produced by *Penicillium patulum*, in the yeast *Saccharomyces cerevisiae*. In order to activate the PKS a 4´-phosphopantetheinyl transferase (PPTase) is required. We therefore co-expressed PPTases encoded by either *sfp* from *Bacillus subtilis* or by *npgA* from *Aspergillus nidulans*. The different strains were grown in batch cultures. Growth and product concentration were measured and kinetic parameters were calculated. It was shown that both PPTases could be efficiently used for activation of PKSs in yeast as good yields of 6-MSA were obtained with both enzymes.

[*] Corresponding auther: ael@biocentrum.dtu.dk

[1] Center for Microbial Biotechnology, BioCentrum-DTU, Denmark

[2] Department of Biology, University of Iceland, Iceland

2.1 Introduction

Polyketides represent a large group of natural products which are found in fungi, bacteria, and plants. This diverse group includes not only compounds that exhibit antibacterial, antifungal, anticancer, and antihelmintic properties, but also cardiovascular agents, cholesterol lowering agents and mycotoxins. Many of them have tremendous economic importance. The antibiotic market amounts to almost US$30 billion including about 160 antibiotics and derivatives such as the macrolide polyketide erythromycin, tetracyclines and others. Other important polyketide products are the hypocholesterolemic statins, immunosuppressants and antitumor compounds, which together have annual sales exceeding US$10 billion (Demain, 2000). For some of these important drugs key patents will expire in the coming years and there will therefore be a shift towards production of these pharmaceuticals as generics, and this is likely to result in increased production volume of many of these compounds.

The biosynthesis of polyketides by polyketide synthases (PKSs) is very similar to the fatty acid biosynthesis with the condensation of acetyl-CoA and malonyl-CoA in a reaction that is catalyzed by ketosynthase (KS). The result of this reaction is an enzyme-bound β-keto thioester. The keto group may be further reduced to a hydroxyl group, an unsaturated thioester; and even a saturated thioester. This leads to a distinct difference from the simple hydrocarbon fatty acid skeleton in that the resulting compounds display a wide structural complexity that is further enhanced by a variety of modifications. PKSs can also use different starter and extender units and hereby further increase the molecular diversity. The structure of a polyketide also depends on the stereochemistry at each chiral center.

Due to the wide bioactivity within this group of compounds the production of novel polyketides has gained much interest in the area of metabolic engineering. Much work has already been accomplished through the use of combinatorial biosynthesis or genetic alteration of postpolyketide pathways as well as using domain shuffling for production of new polyketides using *Streptomyces* (McDaniel et al., 1994; Oliynyk et al., 1996; Menzella et al., 2005). Combinatorial biosynthesis can be obtained through elimination of

unwanted catalytic groups and/or addition of specific moieties or functional groups in the PKS genes.

Fungal PKSs have not been studied to the same extent as bacterial PKSs. Unexploited sources of PKSs are e.g. from lichens, as they are difficult to cultivate. For analysis of fungal PKSs it is of interest to develop a fungal expression system that allows further study of eukaryotic PKSs, but may also potentially be used for high-level expression of polyketides. Yeast is a very suitable system for heterologous expression as it is not only generally regarded as safe (GRAS), but also easy to manipulate genetically, well suited for industrial fermentation and easy to grow and handle in large scale processes. In addition, much knowledge on the genomics and physiology of this organism is available. Although *S. cerevisiae* has no known polyketide biosynthetic pathway, this organism has earlier been used with success as a platform for production of chemicals, and due to the relative ease of performing genetic engineering the metabolic network may readily be improved for enhanced polyketide production. The possibility to expand the range of starter and extender units for polyketide biosynthesis in *S. cerevisiae* has been illustrated by introducing pathways for the production of methylmalonyl-CoA (Mutka et al., 2006).

There has already been one report about successful expression of a fungal polyketide in yeast by Kealey et al., (1998), who demonstrated that it was possible to heterologously produce 6-methylsalicylic acid (6-MSA) in *S. cerevisiae*. However, no detailed physiological characterization was carried out and no details on the production kinetics were given. 6-MSA, which has an eight-carbon chain that cyclizes to form an aromatic ring, is among the smallest polyketides (Katz and Donadio, 1993) and is naturally produced by *Penicillium patulum* as an intermediate in the biosynthesis of the toxin patulin (Staunton and Weissman, 2001). 6-MSA is generally toxic to plants, bacteria and fungi *in vitro* e.g. *Colletotrichum acutatum*, *C. gloeosporioides*, *Botryosphaeria dothidea*, *B. obtuse* and *Peltaster fructicola*. Treatment of apple fruit with 6-MSA reduced infection by *C. acutatum*. (Venkatasubbaiah and Chilton, 1992; Venkatasubbaiah et al., 1995). In solid cultures of *P. patulum*, 6-MSA was produced only when an aerial mycelium was present, indicating that the production is linked to cellular differentiation (Peace, 1981).

The PKS that is responsible for the production of 6-MSA, called 6-MSA synthase (6-MSAS), was the first fungal PKS gene to be cloned and is hence one of the best characterized fungal PKSs (Dimroth et al., 1976; Beck et al., 1990). 6-MSAS catalyzes three successive condensation reactions to form 6-MSA from one molecule of acetyl-CoA and three molecules of malonyl-CoA and uses NADPH as an electron donor (Scott et al., 1971). Besides being expressed in yeast, 6-MSAS from *P. patulum* has been heterologously expressed in *Streptomyces coelicolor* (Bedford et al., 1995), in *Escherichia coli* (Kealey et al., 1998) and in tobacco (Yalpani et. al., 2001).

Production of polyketides in yeast requires co-expression of a 4′-phosphopantetheinyl transferase (PPTase) (Kealey et al., 1998; Andrésson unpublished data). PPTase transfers the 4′-phosphopantetheine (Ppant) moiety from coenzyme A (CoA) to the side chain β-hydroxyl group of a conserved serine residue within the PKS, resulting in the conversion of the Ppant-dependent carrier protein from the inactive apo-form into the active holo-form. PPTases play an essential role in the biosynthesis of polyketides and nonribosomal peptides of which bacteria and fungi are the most distinguished producers.

The objective of this study was to demonstrate high level production of polyketides in yeast and at the same time evaluate the use of two different PPTases for activation of heterologously expressed PKS in yeast; one PPTase from *B. subtilis* and the other from *A. nidulans*. We used 6-MSAS as a model PKS and we performed a physiological characterization of the yeast strains producing 6-MSA in order to evaluate the use of yeast as a general cell factory platform for production of polyketides, e.g. also for expression of PKSs originating from unculturable sources.

2.2 Materials and methods

2.2.1 Plasmids and strains

The yeast strains and plasmids used in this study are listed in Table 2.1.

Table 2.1. List of plasmids and strains used in this work

	Characteristics	Reference
Plasmids		
pRS424	Yeast episomal vector with *TRP1* marker	Christianson et al. (1992)
pRS425	Yeast episomal vector with *LEU2* marker	Christianson et al. (1992)
pUBS1	Carrying 6-MSAS gene from *P. patulum* and identical to a pUC plasmid with a bluescript polylinker	Colin Lazarus [a]
pDKP4832	Carrying *npgA* from *A. nidulans* with *GAL1* promoter, and *URA3* marker	David Keszenman-Pereyra,[b]
p424GAL1	Yeast expression vector with *GAL1* promoter, *CYC1* terminator, and *TRP1* marker	Mumberg et al. (1994)
p425GAL1	Yeast expression vector with *GAL1* promoter, *CYC1* terminator, and *LEU2* marker	Mumberg et al. (1994)
p424sfp	Carrying *sfp* from *B. subtilis* with *GAL1* promoter, *CYC1* terminator, and *TRP1* marker	This study
pGei425U	Carrying *GAL1* promoter, modified multiple cloning site, Kozak sequence and *Kluyveromyces lactis URA3*.	This study
pGei425HSA-PP	Carrying 6-MSAS gene from *P. patulum* with *GAL1* promoter, *CYC1* terminator, and *LEU2* marker	This study
Strains		
CEN.PK 111-27B	MATa;*URA3;HIS3;leu2-3,112;trp1-289;MAL2-8;SUC2*	Peter Kötter [c]
CEN.PK 102-3A	MATa;*ura3-52;HIS3;leu2-3,112;TRP1;MAL2-8;SUC2*	Peter Kötter [c]
IBT100074	CEN.PK 111-27B with empty plasmids pRS424 and pRS425	This study
IBT100075	CEN.PK 111-27B with plasmid pGei425HSA-PP and p424sfp	This study
IBT100076	CEN.PK 102-3A with plasmid pGei425HSA-PP and pDKP4832	This study

[a] School of Biological Sciences, University of Bristol, England

[b] Department of Molecular biology and Biotechnology, University of Sheffield, England

[c] Institut für Mikrobiologie, der Johann Wolfgang Goethe-Universität Frankfurt, Germany

2.2.2 Molecular biology techniques

Transformation of yeast cells was carried out according to Gietz and Woods (2002). *E. coli* DH5α was used for propagation of recombinant plasmids. The preparation of competent *E. coli* and transformation protocol were performed according to Sambrook and Russell (2001).

2.2.3 Plasmid construction

The *P. patulum* 6-MSAS gene (X55776) (Beck et al., 1990) was received from University of Bristol in a pUBS1 plasmid. The pGei425U plasmid carrying *K. lactis URA3* was

originally constructed from the p425GAL1 plasmid (Mumberg et al., 1994). The ends of 6-MSAS gene were amplified by PCR with primers: PPmsaF: 5′–CCCTGCAGGTTA ATTAAGGAGGTACCTTTAAAATGCATTCCGCTGCAACT–3′ and PPm6HSAR: 5′–GGCGCGCCATTTAAATTAATGGTGATGGTGATGATGTTTGGCAAGCTTCTCAG C–3′. The primer PPm6HSAR carried a 6X histidine tag and two additional restriction sites, *Swa*I and *Asc*I and recombination tags for the pGei425U plasmid. The pGei425U plasmid was opened with *Hin*dIII and the pUBS1 plasmid was cut with *Xba*I. The opened plasmid together with the PCR products bridging the ends of the 6-MSAS gene and the flanking plasmid sequences were then transformed into *S. cerevisiae* W303a (MATα; *ade2*;*leu2*;*his*;*trp1*;*ura3*) Recombination removed the *K. lactis* URA3 gene from pGei425U inserting the 6X histidine-tagged 6-MSAS gene yielding pGei425HSA-PP. The construction was confirmed by sequencing.

To construct the p424sfp plasmid, the *sfp* gene of *B. subtilis* strain 1553 (Bacillus genetic stock center) was amplified by PCR with primers EcosfpF: 5'–GCCCGAATTCAGAAT GAAGATTTACGGAAT–3' and ClasfpR: 5'–GCGCATCGATTTATAAAAGCTCTTC GTA–3', cut with *Eco*RI and *Cla*I and ligated into p424GAL1 plasmid that had been opened with *Eco*RI and *Cla*I.

2.2.4 Strain construction

The strain *S. cerevisiae* IBT100074 was constructed by transformation of the plasmid pRS424 and pRS425 into *S. cerevisiae* CEN.PK111-27B. *S. cerevisiae* CEN.PK111-27B was co-transformed with the plasmid carrying the 6-MSAS gene (plasmid pGei425HSA-PP) and the plasmid carrying the *sfp* encoding PPTase from *B. subtilis* (plasmid p424sfp) resulting in *S. cerevisiae* IBT100075. IBT100076 was constructed by co-transforming the plasmid pGei425HSA-PP and the plasmid carrying the *npgA* encoding PPTase from *A. nidulans* (plasmid pDKP4832) into *S. cerevisiae* CEN.PK 102-3A. The strains were selected on synthetic media lacking leucine/tryptophan and leucine/uracil, respectively, in order to obtain the correct transformants.

2.2.5 Bioreactors and cultivation conditions

The strains were cultivated in 100 mL medium in 500 mL baffled shake flasks for propagation. The medium consisted of galactose 10 g/L, $(NH_4)_2SO_4$ 7.5 g/L, KH_2PO_4 14.4 g/L, $MgSO_4.7H_2O$ 0.5 g/L, 0.05 mL/L Antifoam 298 (Sigma-Aldrich, St. Louis, MO, USA) and 2 mL/L trace metal solution ($FeSO_4.7H_2O$ 3 g/L, $ZnSO_4.7H_2O$ 4.5 g/L, $CaCl_2.6H_2O$ 4.5 g/L, $MnCl_2.2H_2O$ 0.84 g/L, $CoCl_2.6H_2O$ 0.3 g/L, $CuSO_4.5H_2O$ 0.3 g/L, $Na_2MoO_4.2H_2O$ 0.4 g/L, H_3BO_3 1 g/L, KI 0.1 g/L, and $Na_2EDTA.2H_2O$ 15 g/L), and 2 mL/L vitamin solution (d-biotin 50 mg/L, Ca-pantothenate 1 g/L, thiamin-HCl 1 g/L, pyridoxin-HCl 1 g/L, nicotinic acid 1 g/L, p-aminobenzoic acid 0.2 g/L, and m-inositol 12.5 g/L). pH of the medium was adjusted to 6.5 by 2M NaOH prior to autoclaving. The cultures were incubated at 30 ^oC, with shaking at 150 rpm (model 3033, GFL, Burgwedel, Germany). When the absorbance (A_{600}) reached 13, 0.1 mL of the culture was used as inoculum for the batch cultivations.

Batch cultivations were carried out in bioreactors (Biostat Braun Biotech International GmbH, Mulsungen, Germany) with a working volume of 2 L. The medium consisted of galactose 20 g/L, $(NH_4)_2SO_4$ 10 g/L, KH_2PO_4 3 g/L, $MgSO_4.7H_2O$ 1 g/L with 0.05 mL/L Antifoam 298 (Sigma-Aldrich, St. Louis, MO, USA), 2 mL/L of trace metal solution and 2 mL/L of vitamin solution. The compositions of the trace metal and vitamin stock solutions were the same as mentioned above. The temperature of the cultivations was kept at 30 ^oC, and the pH was automatically controlled at 5.0 by addition of 2M KOH. The agitation was 500 rpm and the bioreactors were aerated with 2 L air/min.

2.2.6 Absorbance and biomass measurement

Absorbance was measured at 600 nm in a Hitachi U-1100 spectrophotometer throughout the batch cultivations. Biomass dry weight was determined by filtering a known volume of fermentation broth, approximately 10-15 mL depending on the biomass concentration in the media, through a dried, pre-weighed nitrocellulose filter (Gelman Science, Ann Arbor, MI) with a pore size 0.45 μm. The residue was washed twice with distilled water. The

filter was dried to constant weight in a microwave oven at 150 W for 10 minutes, cooled down in a desiccator, and the weight gain was measured.

2.2.7 Analysis of galactose, primary metabolites, and 6-MSA

Fermentation broth was filtered through a 0.45 μm pore-sized cellulose acetate filter (GE Osmonics Labstore, Minnetonka, MN) for later analysis of substrate and products. Galactose and primary metabolites was analyzed by high-performance liquid chromatography (Dionex-HPLC; Sunnyvale, CA), equipped with an Aminex HPX-87H ion exclusion column, 300 mm × 7.8 mm (Bio-Rad Laboratories, Hercules, CA), which was operated at 60 ^{o}C, and a flow rate of 0.6 mL/min of 5 mM H_2SO_4 using a refractive index detector (Shodex RI-71, Tokyo, Japan) and UV detector (UVD340S, Dionex). 6-MSA was quantified from the filtrate by HPLC (Agilent 1100 series) with Luna C18(2) column, 150 × 4.60 mm with 5 micron of the porous silica particles to which the C18 phase is bonded, using a gradient of 1% acetic acid in milliQ water (solvent A) and 1% acetic acid in acetonitrile (solvent B) at a flow rate of 0.5 mL/min. The gradient of the solvents was 20% to 80% B in 30 minutes, then 100% of B for 5 minutes, and 20% B for the next 10 minutes.

2.3 Results

2.3.1 Maximum specific growth rate, biomass yield, and primary metabolites

The fermentation physiology of three different yeast strains, the strain with empty plasmids (IBT100074), the strain containing the plasmids carrying the gene encoding 6-MSAS and PPTase gene from *B. subtilis* (IBT100075) and the strain containing the plasmids carrying 6-MSAS and PPTase gene from *A. nidulans* (IBT100076) were characterized in batch cultivations in a galactose minimal medium. The cultivations were carried out to estimate the kinetic parameters and to investigate the influence of the two different PPTases from *B. subtilis* and *A. nidulans*, respectively, on growth and on production of primary metabolites as well as of 6-MSA. For each strain two cultivations were performed and the mean values were presented. In all cultivations diauxic growth

were observed, and the three strains showed very similar growth characteristics but different growth kinetics (Figure 2.1).

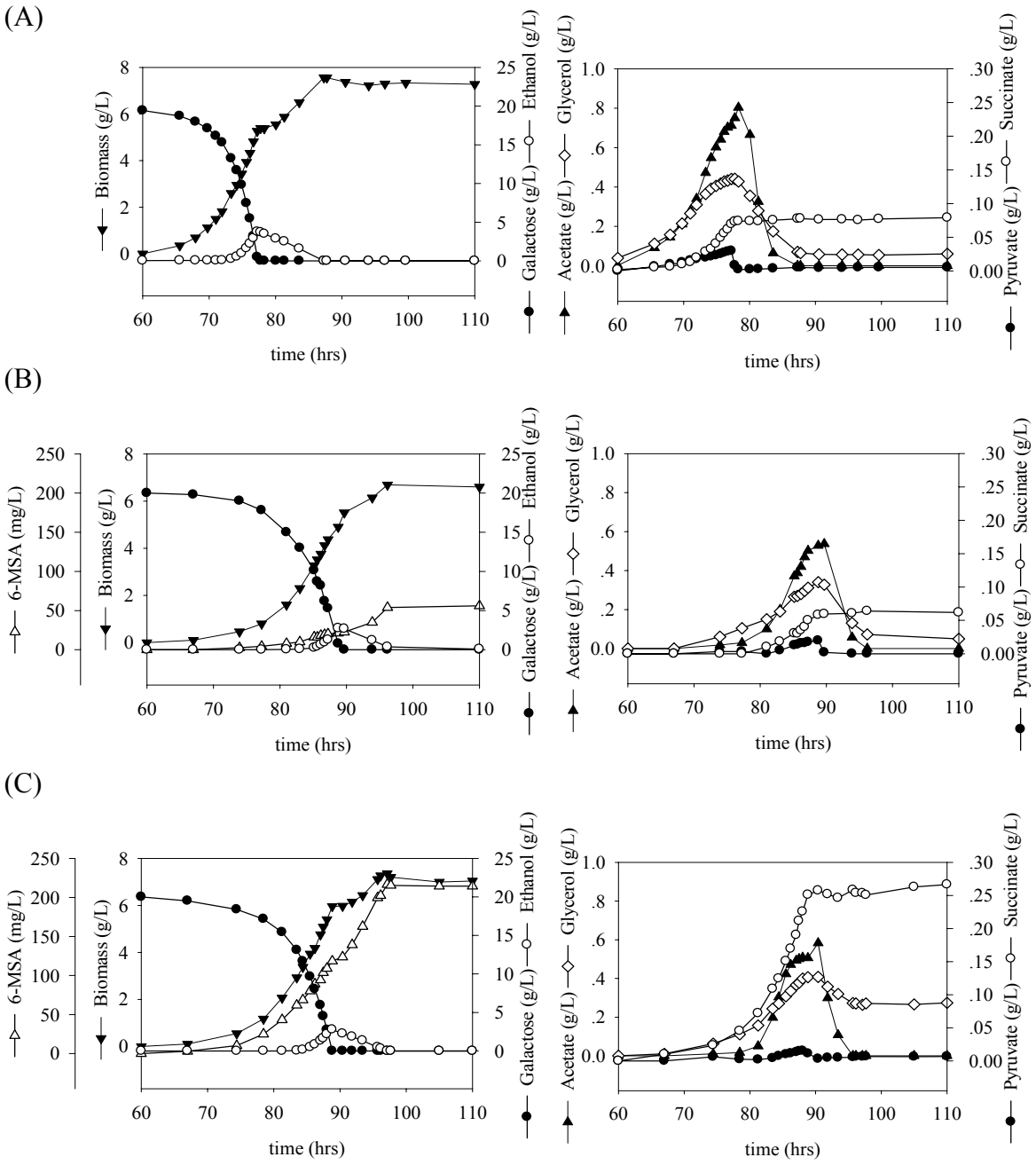


Figure 2.1. Results from batch cultivation with the three different strains. The left panel shows the concentrations of biomass, galactose, ethanol and 6-MSA (except for the strain with the empty plasmids) and the right panel shows the concentrations of the primary metabolites acetate, glycerol, pyruvate and succinate.

A[1]. Results with IBT100074, carrying the empty plasmids pRS424 and pRS425.

B[1]. Results with IBT100075, carrying 6-MSAS gene from *P. patulum* and PPTase gene from *B. subtilis*.

C[1]. Results with IBT100076, carrying 6-MSAS gene from *P. patulum* and PPTase gene from *A. nidulans*.

[1)] The concentration profiles were very similar for the two cultivations. Data shown is for one of the replicates.

After a lag phase, exponential growth was observed for all three strains. During this phase the cells grew at a maximum specific growth rate (μ_{max}) until the galactose was exhausted. The maximum specific growth rate was approximately 0.22 hr^{-1} for the reference strain (IBT100074) and substantially lower for both strains producing 6-MSA, 0.13 and 0.16, respectively (Table 2.2).

Table 2.2. Maximum specific growth rate and product yields during the batch cultures for the three different strains.

Maximum specific growth rate and yields	Strains		
	IBT100074	IBT100075	IBT100076
μ_{max} (hr^{-1})	0.22 ± 0.003	0.13 ± 0.03	0.16 ± 0.006
$Y_{S,X}$ (C-mmol/C-mol gal)	268 ± 8	274 ± 10	294 ± 18
$Y_{S,EtOH}$ (C-mmol/C-mol gal)	348 ± 12	360 ± 10	293 ± 2
$Y_{S,Ace}$ (C-mmol/C-mol gal)	64 ± 1	26 ± 12	18 ± 4
$Y_{S,Pyr}$ (C-mmol/C-mol gal)	0.67 ± 0.16	0.90 ± 0.04	0.89 ± 0.02
$Y_{S,Gly}$ (C-mmol/C-mol gal)	45 ± 3	11 ± 5	52 ± 12
$Y_{S,Suc}$ (C-mmol/C-mol gal)	5.7 ± 0.4	1.4 ± 0.4	16.5 ± 3.4
$Y_{S,CO2}$ (C-mmol/C-mol gal)	278 ± 0.7	324 ± 54	324 ± 13
$Y_{S,6\text{-}MSA}$ (C-mmol/C-mol gal)	0	1.6 ± 0.6	8.6 ± 0.3
$Y_{EtOH,6\text{-}MSA}$ (C-mmol/C-mol EtOH)	0	14 ± 2	42 ± 9
r_p (mg 6-MSA/g dw/hr) in gal phase	0	0.6 ± 0.1	2.9 ± 0.4
r_p (mg 6-MSA/g dw/hr) in EtOH phase	0	0.8 ± 0.1	2.9 ± 0.2
C-blannce (%)	100.9	99.9	100.7

In the first exponential growth phase there is over-flow metabolism resulting in formation of ethanol (Figure 2.1). Other primary metabolites such as acetate, succinate, glycerol and pyruvate were also excreted to the media in this phase but in much smaller amounts. The yields of the different primary metabolites produced during the first exponential phase were calculated and the results are collected in Table 2.2. The biomass composition was calculated based on formula $CH_{1.8}O_{0.5}N_{0.2}$ (Roels, 1983).

For all three cultivations the consumed galactose carbon could be accounted for in the product with an error margin of 1%. When galactose was depleted there was a second exponential phase. In this phase acetate, glycerol and pyruvate formed during the first exponential growth phase were successively consumed, and eventually the cells started to consume ethanol resulting in a second exponential phase where biomass and CO_2 were produced. When all the ethanol was consumed, the CO_2 production rapidly decreased to zero marking the end of the second exponential growth phase. The CO_2 from the exhaust gas during one of the batch cultivations is shown in Figure 2.2. After the second exponential phase the cultures reached the stationary phase where no more biomass or metabolic products were formed.

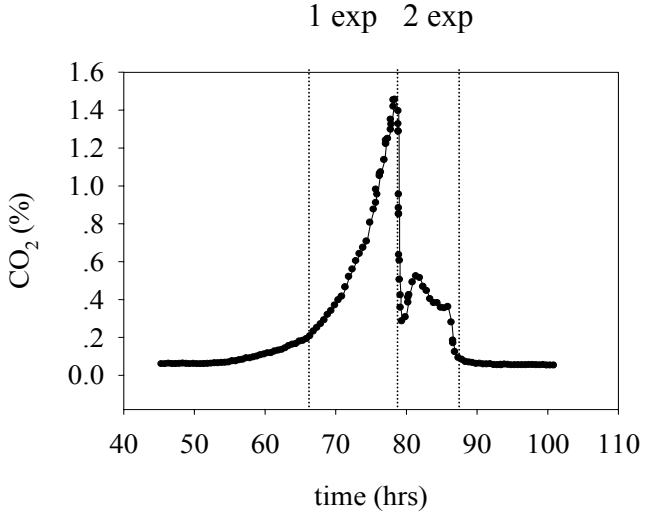


Figure 2.2. CO_2 in the exhaust gas during a batch cultivation of IBT100076. The CO_2 profile clearly illustrates the first and the second exponential growth phase of the cultivation.

2.3.2 6-MSA production and yield

From the measurements of the 6-MSA concentration, the biomass concentration and the specific growth rate it was possible to calculate both the 6-MSA yield and the specific production rate of 6-MSA (Table 2.2). Hereby it is possible to evaluate the growth kinetics for the two different growth phases and also the influence of the PPTase. As expected, there was no production of 6-MSA by the reference strain with the empty plasmids (IBT100074), whereas for the two other strains, IBT100075 and IBT100076, there was 6-MSA production simultaneously with growth.

During the first exponential phase where galactose was consumed, the yield of 6-MSA was 8.6 C-mmol/C-mol galactose for the strain with PPTase gene from *A. nidulans* (IBT100076), whereas the strain with PPTase gene from *B. subtilis* produced 1.6 C-mmol 6-MSA/C-mol galactose (Table 2.2). In the second exponential growth phase where the cells used ethanol as a carbon source, 6-MSA was produced with yields of 42 and 14 C-mmol 6-MSA/C-mol ethanol, from the strains expressing PPTase from *A. nidulans* and *B. subtilis*, respectively. After the ethanol was consumed, the cultures reached a stationary phase and here the final titer of 6-MSA was above 200 mg/L for the strain expressing the PPTase from *A. nidulans* (Figure 2.1).

2.3.3 Bioinformatics of the PPTases

In order to obtain some insight into the possible molecular mechanisms resulting in the difference in productivity arising from using the two different PPTases, we compared the two proteins using bioinformatics. The *npgA* gene (GenBank accession number, AAF12814) is predicted to encode a peptide of 344 amino acids (aa), and is more than twice the size of the Sfp peptide (AAC36829) which is only 165 aa in length. A BlastP alignment (Altschul et al., 1997) of the two peptides reveals only 23% (46/195) identity between the sequences and although this similarity is quite weak, the identical aa correspond to conserved residues previously determined, by a combination of crystallization and sequence alignment, to be vital domains in Sfp (Reuter et al., 1999). Alignment of PPTase homologues from several species confirms that the residues are

generally conserved (Keszenman-Pereyra et al., 2003). Since the N-terminal part of NpgA seems to have the same domains as Sfp, we took a closer look at the final 100 aa of the C-terminal in NpgA which are unique when aligned to Sfp. A multiple sequence alignment of this region of NpgA with four fungal PPTase homologues (*A. fumigatus* (EAL93155), *A. oryzae* (BAE65284), *A. terreus* (XP_001218317), *A. niger* (Joint Genome Institute, ASN211634)) using ClustalX 1.8 (Jeanmougin et al., 1998) showed that the N-terminal of these aa sequences are indeed conserved (30%), and contained the conserved motif "A291 TAARG". In light of this, it is possible that the conserved N-terminal mediates a specific function that is not present in Sfp and that this additional function accounts for the improved productivity of 6-MSA in our experiments. Additionally, NpgA is phylogenetically closer to the *P. patulum* PKS than the bacterial Sfp protein. Thus, although both Sfp and NpgA can be described as broad-specificity PPTases, the *Aspergillus* protein seems to be slightly better adapted for activation of fungal PKSs.

2.4 Discussion

In the present study, we demonstrated the heterologous expression of a PKS in *S. cerevisiae*. There are several reasons for the expression of polyketide pathways from a natural host in a heterologous host: 1) many useful polyketides originate from a source that is difficult or even impossible to cultivate, 2) heterologous expression provides an efficient platform for combinatorial biosynthesis, and 3) by using a common cell factory platform for production of polyketides, one may optimize the central carbon metabolism for efficient production of the precursors required for polyketide biosynthesis, and hereby the same cell factory platform may be used for high level expression of several polyketides. A fungal polyketide, 6-MSA produced from *P. patulum*, was chosen as a model polyketide in this study due to its simple structure, as it has been well studied, and it has been previously established in yeast (Kealey et al., 1998).

In our study we showed that it was possible to produce large amounts of 6-MSA in yeast, and we obtained 6-MSA yields as high as 18 mg 6-MSA/g DW corresponding to 200 mg of 6-MSA/L with the PPTase from *A. nidulans*. This yield was superior compared to the

natural producers *P. griseofulvum* IBT3115 and *A. terreus* IBT12713 which produced 0.2 and 0.1 mg of 6-MSA/L in minimal media (Wattanachaisaereekul et al.,unpublished data), respectively. Furthermore, only 0.58 mg of 6-MSA/plate was produced by *P. patulum* growing on solid glucose-based Czapek-Dox agar (medium volume 40 mL) at 90 hrs after inoculation (Peace et al., 1981). The yield from this study is comparable to yields obtained for endogeneously produced nystatin by *Streptomyces noursei* (Jonsbu et al., 2001; Nielsen and Eliasson, 2004), which is quite impressive and indicates that it is possible to produce even more complex polyketides at reasonable production levels. In terms of strain stability the strains constructed and used in this study were found to be stable throughout the batch cultivations, as clearly indicated by the constant specific productivity in the different growth phases.

In the process of developing our yeast producing polyketide platform we evaluated two different PPTases. This class of enzyme has been classified into three groups according to their size and substrate selectivity (Lambalot et al., 1996). The first group is the AcpS-type referred to as bacterial holo-acyl CP synthases. PPTases of this class are characterized by a size of 120-140 amino acids and a restricted substrate tolerance. Many organisms have an AcpS-type PPTase for the modification of fatty acid synthase acyl carrier protein such as PPT2 in *S. cerevisiae* and AcpS in *E. coli* and *B. subtilis*. The second group of PPTases comprises enzymes that are integral parts of eukaryotic FAS, e.g. the yeast PPTase found in FAS2. The last group, sfp-type PPTases, contains enzymes that are about twice the size of the AcpS-type PPTases (about 240 amino acids). PPTases in this group act as monomers have a broad substrate spectrum and are usually closely connected to secondary metabolism (Finking and Marahiel, 2004; Mofid et al., 2004; Gross et al., 2005). The best-characterized representative of this class is the sfp-type PPTase from *B. subtilis*. Additional members of sfp-type PPTases include PPTases of higher organisms, e.g. human PPTase (Joshi et al., 2003), Lys5 of *S. cerevisiae* which is responsible for the modification of multifunctional Lys2 (Ehmann et al., 1999; Mootz et al., 2002), and NpgA from *A. nidulans* required for the modification of at least five different enzymes involved in both primary and secondary metabolism such as pigment synthesis, lysine biosynthesis,

and the biosynthesis of siderophores and penicillin (Mayorga and Timberlake, 1992; Oberegger et al., 2003; Keszenman-Pereyra et al., 2003).

Since the PPTases identified in *S. cerevisiae* are not involved in the production of any secondary metabolites and do not activate the *P. patulum* 6-MSAS (Kealey et al., 1998; Andrésson unpublished data), we expressed two different sfp-type PPTases, Sfp from *B. subtilis* and NgpA from *A. nidulans*. Both enzymes are known to show broad substrate specificity, and we were therefore interested in investigating the production kinetics when these two different PPTases were used. The results from the batch cultivations showed that during the first exponential phase the strain carrying the PPTase from *A. nidulans* (IBT100076) produced higher amounts of all primary metabolites than the strain expressing the PPTase from *B. subtilis* (IBT100075) as shown in Figure 2.1. It is interesting that the increased production of primary metabolites also resulted in a substantially higher 6-MSA production. As observed in the second exponential phase, where the ethanol formed during the first exponential phase was consumed, the yield of 6-MSA from IBT100076 was 42 C-mmol/C-mol ethanol, whereas IBT100075 yielded 14 C-mmol 6-MSA/C-mol ethanol.

When the yield of 6-MSA on galactose was compared to the yield of 6-MSA on ethanol, we found that the use of ethanol as carbon source resulted in substantially more 6-MSA than when galactose was the carbon source. This is likely due to the fact that the use of a C2 carbon source may result in a higher level of the precursor (acetyl-CoA) in the cytosol (Seker et al., 2005; van den Berg et al., 1996, 1998) for production of 6-MSA. Furthermore, during the conversion of ethanol into acetaldehyde and further into acetate (that can be converted into acetyl-CoA) NADPH is generated, and NADPH is required as an electron donor for production of polyketides in the reduction of the keto group to a hydroxyl group or further. Thus, for production of polyketides by yeast it is beneficial to use ethanol as carbon source.

In summary we have demonstrated that PPTases from both *B. subtilis* and *A. nidulans* can be used for the activation of PKSs in *S. cerevisiae*, but expression of the PPTase from *A.*

nidulans leads to higher production of the polyketide. Furthermore, through the use of ethanol as the carbon source the production is more efficient than with galactose (and likely other sugars). We reached more than 200 mg/L of 6-MSA in a batch fermentation and yields comparable with that obtained for *Streptomyces* producing endogeneous antibiotics.

2.5 References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Beck J, Ripka S, Siegner A, Schiltz E, Schweizer E. 1990. The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*: Its gene structure relative to that of other polyketide synthases. Eur. J. Biochem. 192: 487-498.

Bedford DJ, Schweizer E, Hopwood DA, Khosla C. 1995. Expression of a functional polyketide synthase in the bacterium *Streptomyces coelicolor* A3(2). J. Bacteriol. 177(15): 4544-4548.

Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. 1992. Multifunctional yeast high-copy-number shuttle vectors. Gene 110: 119-122.

Demain AL, 2000. Small bugs, big business: The economic power of the microbe. Biotechnology Advances 18: 499-514.

Dimroth P, Ringelmann E, Lynen F. 1976. 6-Methylsalicylic acid synthetase from *Penicillium patulum*. Some catalytic properties of the enzyme and its relation to fatty acid synthetase. Eur. J. Biochem. 68: 591-596.

Ehmann DE, Gehring AM, Walsh CT. 1999. Lysine biosynthesis in *Saccharomyces cerevisiae*: mechanism of α–aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. Biochemistry 38: 6171-6177.

Finking R, Marahiel MA. 2004. Biosynthesis of nonribosomal peptides. Annu. Rev. Microbiol. 58: 453-488.

Gietz RD, Woods RA. 2002. Transformation of yeast by the LiAc/SS carrier DNA/PEG method. Methods in Enzymology 350: 87-96.

Gross F, Gottschalk D, Müller R. 2005. Posttranslational modification of myxobacterial carrier protein domains of *Pseudomonas* sp. by intrinsic phosphopantetheinyl transferase. Appl. Microbiol. Biotechnol. 68: 66-74.

Jeanmougin F., Thompson JD, Gouy M, Higgins DG, Gibson TJ. 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. 23: 403–405.

Jonsbu E, Christensen B, Nielsen J. 2001. Changes of in vivo fluxes through central metabolic pathways during the production of nystatin by *Streptomyces noursei* in batch culture. Appl. Microbiol. Biotechnol. 56: 93-100.

Joshi AK, Zhang L, Rangan VS, Smith S. 2003. Cloning, expression, and characterization of a human 4′-phosphopantetheinyl transferase with broad substrate specificity. J. Biol. Chem. 278(35): 33142-33149.

Katz L, Donadio S. 1993. Polyketide synthesis: Prospects for hybrid antibiotics. Annu. Rev. Microbiol. 47: 875-912.

Kealey JT, Liu L, Santi DV, Betlach MC, Barr PJ. 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proc. Natl. Acad. Sci. 95: 505-509.

Keszenman-Pereyra D, Lawrence S, Twfieg ME, Price J, Turner G. 2003. The *npgA*/*cfwA* gene encodes a putative 4′-phosphopantetheinyl transferase which is essential for penicillin biosynthesis in *Aspergillus nidulans*. Curr. Gent. 43: 186-190.

Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT. 1996. A new enzyme superfamily: the phosphopantetheinyl transferases. Chem. Biol. 3: 923-936.

Mayorga ME, Timberlake WE. 1992. The developmentally regulated *Aspergillus nidulans wA* gene encodes a polypeptide homologous to polyketide and fatty acid synthases. Mol. Gen. Genet. 235: 205-212.

McDaniel R, Ebert-Khosla S, Fu H, Hopwood DA, Khosla C. 1994. Engineered biosynthesis of novel polyketides: influence of a downstream enzyme on the catalytic specificity of a minimal aromatic polyketide synthase. Proc. Natl. Acad. Sci. USA. 91: 11542-11546.

Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, Patel KG, Hopwood DA, Santi DV. 2005. Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. Nature Biotechnology 23(9): 1171-1176.

Mofid MR, Finking R, Essen LO, Marahiel MA. 2004. Structure based mutational analysis of the 4′-phosphopantetheinyl transferases Sfp from *Bacillus subtilis*: Carrier protein recognition and reaction mechanism. Biochemistry 43: 4128-4136.

Mootz HD, Schörgendorfer K, Marahiel MA. 2002. Functional characterization of 4′-phosphopantetheinyl transferase genes of bacterial and fungal origin by complementation of *Saccharomyces cerevisiae* Lys5. FEMS Microbiol lett. 213(1): 51-57.

Mumberg D, Müller R, Funk M. 1994. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Research. 22(25): 5767-5769.

Mutka SC, Bondi SM, Carney JR, Da Silva NA, Kealey JT. 2006. Metabolic pathway engineering for complex polyketide biosynthesis in *Saccharomyces cerevisiae*. FEMS Yeast Res. 6: 40-47.

Nielsen J, Eliasson A. 2004. From Glucose to Antibiotics: What controls the Fluxes? *In* Wohlleben W, Spelling T, Mueller-Tiemann B (eds.). Ernst Schering research foundation workshop, band51: Biocombinatorial approaches for drug finding, Springer-Verlag, Heidelberg, Germany, Chapter 9: 195-214.

Oberegger H, Eisendle M, Schrettl M, Graessle S, Haas H. 2003. 4´-phosphopantetheinyl transferase-encoding *npgA* is essential for siderophore biosynthesis in *Aspergillus nidulans*. Curr. Genet. 44: 211-215.

Oliynyk M, Brown MJB, Cotes J, Staunton J, Leadlay PF. 1996. A hybrid modular polyketide synthase obtained by domail swapping. Chem Biol 3: 833-839.

Peace JN, Bartman CD, Doerfler DL, Campbell IM. 1981. 6-Methylsalicylic acid production in solid cultures of *Penicillium patulum* occurs only when an aerial mycelium is present. Applied and Environmental Microbiology. 41(6): 1407-1412.

Reuter K, Mofid MR, Marahiel MA, Ficner R. 1999. Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4´-phosphopantetheinyl transferase superfamily. EMBO J. 18: 6823-6831.

Roels JA. 1983. Energetics and kinetics in biotechnology. Amsterdam: Elsevier.

Sambrook J, Russell DW. 2001. Molecular Cloning: A laboratory manual (3rd editon). Cold Spring Harbor Laboratory Press, NY, USA.

Scott AI, Phillips GT, Kircheis U. 1971. Biosynthesis of polyketides: The synthesis of 6-methylsalicylic acid and triacetic acid lactone in *Penicillium patulum*. Bioorg. Chem. 1: 380-399.

Seker T, Møller K, Nielsen J. 2005. Analysis of acyl CoA ester intermediates of the mevalonate pathway in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol. 67: 119-224.

Staunton J, Weissman KJ. 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18: 380-416.

van den Berg MA, Jong-Gubbels P de, Kortland CJ, Dijken JP van, Pronk JT, Steensma HY. 1996. The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation. J. Biol. Chem. 271: 28953-28959.

van den Berg MA, Jong-Gubbels P de, Steensma HY. 1998. Transient mRNA responses in chemostat cultures as a method of defining putative regulatory elements: application to genes involved in *Saccharomyces cerevisiae* acetyl-coenzyme A metabolism. Yeast 14: 1089-1104.

Venkatasubbaiah P, Chilton WS. 1992. An epoxydon-derived ester from a *Phoma* sp. pathogenic to rhubarb. J. Nat. Prod. 55: 639–643.

Venkatasubbaiah P, Sutton TB, Chilton WS. 1995. The structure and biological properties of secondary metabolites produced by *Peltaster fructicola*, a fungus associated with apple sooty blotch disease. Plant Dis. 79: 1157-1160.

Yalpani N, Altier DJ, Barbour E, Cigan AL, Scelonge CJ. 2001. Production of 6-methylsalicylic acid by expression of a fungal polyketide synthase activates disease resistance in tobacco. The Plant Cell. 13: 1401-1409.

Chapter 3

Production of the Polyketide 6-MSA in Yeast Engineered for Increased Malonyl-CoA Supply

Songsak Wattanachaisaereekul, Anna Eliasson Lantz*, Michael Lynge Nielsen, and Jens Nielsen

Manuscript submitted to Metabolic Engineering (2007)

Abstract

The heterologous production of fungal polyketides was investigated using 6-methylsalicylic acid synthase (6-MSAS) as a model polyketide synthase and *Saccharomyces cerevisiae* as a host. In order to improve the production of 6-MSA by enhancing the supply of precursors, the promoter of the gene (*ACC1*) encoding acetyl-CoA carboxylase, which catalyzes the conversion of acetyl-CoA to malonyl-CoA, was replaced with a strong, constitutive promoter (*TEF1*p) in a strain harboring two plasmids carrying the genes encoding 6-MSAS from *P. patulum* and PPTase from *Aspergillus nidulans*, respectively. The strain was characterized in batch cultivations with a glucose minimal media (20 g/L), and a 60% increase in 6-MSA titer was observed compared to a strain having the native promoter in front of *ACC1*. The production of 6-MSA was scaled up by the cultivation of the *ACC1-TEF1* promoter strain in minimal media containing 50g/L of glucose, and hereby a final titer of 600 mg/L of 6-MSA was obtained.

Keywords: acetyl-CoA carboxylase, malonyl-CoA, polyketides, heterologous expression, precursor supply

* Corresponding auther: ael@biocentrum.dtu.dk

3.1 Introduction

Polyketides, which are produced through the successive condensation of small carboxylic acids, represents a large group of secondary metabolites with a broad range of structures and biological activities (Katz and Donadio, 1993). The major sources of these compounds are microbial secondary metabolites. Fungi are very potent producers of polyketide metabolites and they are often targets in searches for bioactive polyketides that can be used in pharmaceutical and agrochemical industries. Examples of such bioactive compounds are the antibiotic colletodiol (tetraketide) from *Cytospora* sp. (O'Neill et al., 1993), inhibitors for cholesterol biosynthesis such as compactin (nonaketide) and squalestatin (hexaketide), from *Penicillium citrinum* and *Phoma* sp., respectively (O'Hagen, 1991; Dawson et al., 1992), and the antifungal monocerin (heptaketide) from *Drechslera monoceras* (Turner and Aldridge, 1970).

A wide range of structures of fungal polyketides, several of which have unusual features that are not found among bacterial metabolites, indicates an enzymological diversity which could be exploited in the generation of novel polyketides via combinatorial biosynthesis and metabolic engineering. The potential of metabolic engineering of fungal polyketides as well as the industrial production of useful polyketides will be greatly enhanced if the fungal genes can be heterologously expressed in a well-characterized host such as the yeast *Saccharomyces cerevisiae*. Unlike filamentous fungi, *S. cerevisiae* is a unicellular organism which is easy to work with in terms of large scale fermentation, and as molecular biological techniques are far more advanced for *S. cerevisiae* than for filamentous fungi, *S. cerevisiae* is a very attractive cell factory for heterologous production of fungal polyketides. There are, however, some challenges for the heterologous production of fungal polyketides in *S. cerevisiae*, e.g. the requirement for post-translational modification of the synthases by phosphopantetheinylation, but this has been solved and it has been demonstrated possible to heterologously produce the polyketide, 6-methylsalicylic acid (6-MSA), in *S. cerevisiae* (Kealey et al., 1998). More recently, *S. cerevisiae* carrying the 6-MSAS gene from *P. patulum* has been elaborately characterized in batch cultivations using galactose minimal media and a good yield of 6-

MSA was obtained by the co-expression of the gene *npgA* encoding the enzyme phosphopantetheinyl transferase (PPTase) from *A. nidulans* (Wattanachaisaereekul et al., 2007).

Despite the success in the above mentioned stories we were, however, interested to further increase the productivity and titer of 6-MSA in order to evaluate whether *S. cerevisiae* could serve as a platform for heterologous production of fungal polyketides. We therefore applied metabolic engineering to increase the supply of precursors required for polyketide production. Furthermore, to circumvent the problem of slow growth on galactose we replaced the expression promoter for the polyketide synthase with an endogeneous glucose promoter such that the polyketide 6-MSA could be produced during growth on glucose. As in our previous study the model polyketide used is 6-MSA. To increase the precursor supply we overexpressed the gene encoding the enzyme acetyl-CoA carboxylase (*ACC1*), which catalyzes the carboxylation of acetyl-CoA to form malonyl-CoA. The over-expression of *ACC1* was done by replacing the native promoter of *ACC1* with the *TEF1* promoter, which is considered as a strong, constitutive promoter. We then performed a physiological characterization of the *ACC1* promoter recombinant strain, the native *ACC1* promoter strain, and a non-producing reference strain in order to build a solid platform for the production of polyketides in *S. cerevisiae*.

3.2 Materials and methods

3.2.1 Plasmids and strains

The plasmids and stains used in this study are listed in Table 3.1.

Table 3.1. List of plasmids and strains

	Characteristics	Reference
Plasmids		
pRS424	Yeast episomal vector with *TRP1* marker	Christianson et al. (1992)
pRS426	Yeast episomal vector with *URA3* marker	Christianson et al. (1992)
pDKP4832	Carrying *npgA* from *A. nidulans* with *GAL1* promoter, and *URA3* marker	David Keszenman-Pereyra [a]
pWJ1042	Plasmid containing *Kl URA3* flanked by direct repeats	In our laboratory
pRS306CRUDT	Integrating vector carrying *CYC1* terminator, direct repeat, Up-Down sequence of integration site – *YMLCΔ2*, *TEF1* promoter, and *URA3* marker	Chapter 4
pRS306CRUDTMSA-PP	Integrating vector carrying *CYC1* terminator, direct repeats, Up-Down sequence of integration site – *YMLCΔ2*, *TEF1* promoter, *URA3* marker and 6-MSAS gene from *P. patulum*	Chapter 4
pRS426CT	Expression vector containing *TEF1* promoter, *CYC1* terminator, and *URA3* marker	This study
pRS424CT	Expression vector containing *TEF1* promoter, *CYC1* terminator, and *TRP1* marker	This study
pRS426CTMSA-PP	Expression vector containing *TEF1* promoter, *CYC1* terminator, *URA3* marker, and 6-MSAS gene from *P. patulum*	This study
pRS424CTnpgA	Expression vector containing *TEF1* promoter, *CYC1* terminator, *TRP1* marker, and *npgA* from *A. nidulans*	This study
Strains		
CEN.PK113-9D	MATa;*ura3-52;HIS3;LEU2;trp1-289;MAL2-8;SUC2*	Peter Kötter [b]
CEN.PK ACC	CEN.PK 113-9D which promoter of *ACC1* has been changed to *TEF1* promoter	This study
IBT100081	CEN.PK 113-9D with empty plasmids pRS426 and pRS424	This study
IBT100082	CEN.PK 113-9D with plasmids pRS426CTMSA-PP and pRS424CTnpgA	This study
IBT100083	CEN.PK ACC with plasmids pRS426CTMSA-PP and pRS424CTnpgA	This study

[a] Department of Molecular biology and Biotechnology, University of Sheffield, England

[b] Institut für Mikrobiologie, der Johann Wolfgang Goethe-Universität Frankfurt, Germany

Oligonucleotides were synthesized by MWG-Biotech AZ, Germany (Table 3.2). PCR reactions were performed using Phusion polymerase (Finnzymes Oy, Espoo, Finland) according to the manufacturer's guideline. GFX spin columns from Amersham Biosciences (Buckinghamshire, UK) were used for purification of DNA from either gel band or solution. Total genomic DNA extraction from *S. cerevisiae* for PCR template, DNA ligation, restriction endonuclease digestion and analysis were done in accordance with standard protocols (Sambrook and Russell, 2001)

Transformation of yeast cells was carried out according to Gietz and Woods (2002). *E. coli* DH5α was used for propagation of recombinant plasmids. The preparation of competent *E. coli* and transformation protocol were performed according to Sambrook and Russell (2001).

Table 3.2. List of primers used in this study. The underline sections indicate the restriction sites introduced to facilitate cloning.

DNA fragment	Primer	Sequence 5´- 3´
npgA of *A. nidulans*	SWA1	GGACTAGTAAGCTTGGTACCTTCCACAAGCCT
	SWA2	CCATCGATATTGGACTCATATACCCTGCGAGC
Upstream of *ACC1* promoter	SWA3	CACAATTGTTATCGGTTCTAC
	SWA4	GCAGGGATGCGGCCGCTGACCTTGCTCTGAATCTGAATTCC
Direct repeat1 and *Kl. URA3*	SWA5	GTCAGCGGCCGCATCCCTGCTTCGGCTTCATGGCAA
	SWA6	GAGCAATGAACCCAATAACGAAATC
Direct repeat 2 and *Kl. URA3*	SWA7	CTTGACGTTCGTTCGACTGATGAGC
	SWA8	CACGGCGCGCCTAGCAGCGGTAACGCCAGGGTTTTC
TEF1 promoter	SWA9	CGCTGCTAGGCGCGCCGTGCACACACCATAGCTTCAAAATGTT
	SWA10	AATAAGCTTTCTTCGCTCATTTTGTAATTAAAACTTAGATTAGA
Downstream of *ACC1* promoter	SWA11	ATGAGCGAAGAAAGCTTATTCGAGTCTTCTCCACAGAAGATGGA
	SWA12	TCTCGGAGGCGTGACCCCAG

3.2.2 Bioreactors and cultivation conditions

The strains were cultivated in 100 mL medium in 500 mL baffled shake flasks for propagation. The medium consisted of glucose 10 g/L, $(NH_4)_2SO_4$ 7.5 g/L, KH_2PO_4 14.4 g/L, $MgSO_4.7H_2O$ 0.5 g/L, 0.05 mL/L Antifoam 298 (Sigma-Aldrich, St. Louis, MO, USA) and 2 mL/L trace metal solution ($FeSO_4.7H_2O$ 3 g/L, $ZnSO_4.7H_2O$ 4.5 g/L, $CaCl_2.6H_2O$ 4.5 g/L, $MnCl_2.2H_2O$ 0.84 g/L, $CoCl_2.6H_2O$ 0.3 g/L, $CuSO_4.5H_2O$ 0.3 g/L, $Na_2MoO_4.2H_2O$ 0.4 g/L, H_3BO_3 1 g/L, KI 0.1 g/L, and $Na_2EDTA.2H_2O$ 15 g/L), and 2 mL/L vitamin solution (d-biotin 50 mg/L, Ca-pantothenate 1 g/L, thiamin-HCl 1 g/L, pyridoxin-HCl 1 g/L, nicotinic acid 1 g/L, p-aminobenzoic acid 0.2 g/L, and m-inositol 12.5 g/L). pH of the medium was adjusted to 6.5 by 2M NaOH prior to autoclaving. The cultures were incubated at 30 ^oC, with shaking at 150 rpm. When the absorbance (A_{600})

reached 13 (2.3 g biomass/L), 0.1 mL of the culture was used as inoculum for the batch cultivations.

Batch cultivations were carried out in bioreactors (Biostat Braun Biotech International GmbH, Mulsungen, Germany) with a working volume of 2 L. The medium consisted of glucose 20 g/L or 50 g/L depending on the experiment, $(NH_4)_2SO_4$ 10 g/L, KH_2PO_4 3 g/L, $MgSO_4.7H_2O$ 1 g/L with 0.05 mL/L Antifoam 298, 2 mL/L of trace metal solution and 2 mL/L of vitamin solution. The compositions of the trace metal and vitamin stock solutions were the same as mentioned above. The temperature of the cultivations was kept at 30 ^oC, and the pH was automatically controlled at 5.0 by addition of 2M KOH. The agitation was 500 rpm and the bioreactors were aerated with 2 L air/min.

3.2.3 Biomass measurement

Biomass dry weight was determined by filtering a known volume of fermentation broth, approximately 10-15 mL depending on the biomass concentration in the medium, through a dried, pre-weighed nitrocellulose filter (Gelman Science, Ann Arbor, MI) with a pore size of 0.45 μm. The residue was washed twice with distilled water. The filter was dried to constant weight in a microwave oven at 150 W for 10 minutes, cooled down in a desiccator, and the weight gain was measured.

3.2.4 Analysis of glucose, primary metabolites and 6-MSA

Fermentation samples were filtered through a 0.45 μm pore-sized cellulose acetate filter (GE Osmonics Labstore, Minnetonka, MN) for later analysis of substrate and products. Glucose and primary metabolites were analyzed by high-performance liquid chromatography (Dionex-HPLC; Sunnyvale, CA), equipped with an Aminex HPX-87H ion exclusion column, 300 mm × 7.8 mm (Bio-Rad Laboratories, Hercules, CA), which was operated at 60 ^oC, and a flow rate of 0.6 mL/min of 5 mM H_2SO_4 using a refractive index detector (Shodex RI-71, Tokyo, Japan) and UV detector (UVD340S, Dionex, Softron GmbH, Germering, Germany). 6-MSA was quantified from the filtrate by HPLC (Agilent 1100 series) with a Luna C18(2) column, 150 × 4.60 mm with 5 micron of the

porous silica particles to which the C18 phase is bonded, using a gradient of 50 ppm trifluoroacetic acid (TFA) in milliQ water (solvent A) and 50 ppm TFA in acetonitrile (solvent B) at a flow rate of 1 mL/min. The gradient of the solvents was 20% to 60% B in 10 minutes, then 20% B for the next 2 minutes.

3.3 Results

The results are divided into three parts. The first part describes the construction of the plasmids and strains. The second part describes characterization in batch cultivations of the three different strains, namely the reference strain (IBT100081) carrying the empty plasmids, the *ACC1*-native promoter strain (IBT100082) and the *ACC1-TEF1* promoter strain (IBT100083). The last two strains were carrying the 6-MSAS gene from *P. patulum* and the *npgA* gene encoding the enzyme phosphopantetheinyl transferase from *A. nidulans* under expression of the *TEF1* promoter. The third part of the result section describes the influence of different glucose concentration (20 and 50 g/L) on the physiology and 6-MSA production of the *S. cerevisiae ACC1-TEF1* promoter strain (IBT100083).

3.3.1 Plasmid construction

The *TEF1* promoter and *CYC1* terminator fragment was released together from pRS306CRUDT by *Kpn*I and *Not*I, and ligated into *Kpn*I/*Not*I sites of pRS424 and pRS426 to yield the expression vectors pRS424CT and pRS426CT, respectively. The plasmid pRS426CT was opened by *Not*I and *Cla*I and ligated to the fragment of the 6-MSAS gene which was obtained from plasmid pRS306CRUDTMSA-PP by *Not*I and *Cla*I. The resulting plasmid was designated as plasmid pRS426CTMSA-PP. To construct pRS424CTnpgA, the *npgA* gene which encodes the enzyme phosphopantetheinyl transferase was amplified from pDKP4832 with primers SWA1 and 2. After digestion of the PCR product by *Spe*I and *Cla*I, it was ligated into pRS424CT which had been cut with the same restriction enzymes, *Spe*I and *Cla*I (Figure 3.1)

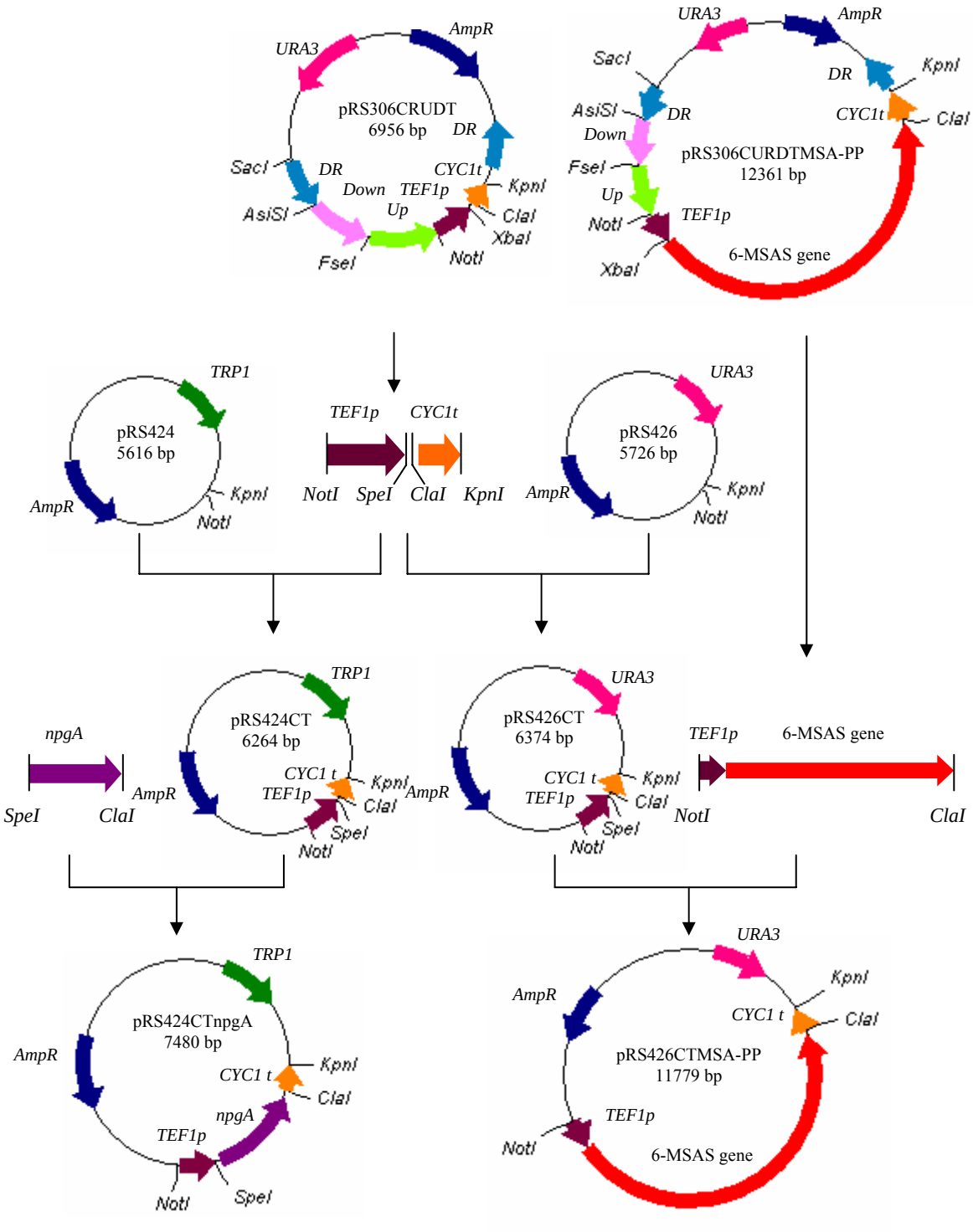


Figure 3.1. Diagram illustrating the construction of plasmids.

3.3.2 Strain construction

S. cerevisiae IBT100081 was constructed by transformation of the empty plasmids pRS424 and pRS426 into *S. cerevisiae* CEN.PK 113-9D. The plasmid carrying the 6-MSA gene from *P. patulum* (pRS426CTMSA-PP) and the plasmid carrying *npgA* encoding PPTase from *A. nidulans* (pRS424CTnpgA) which both have the genes under the control of *TEF1* promoter were co-transformed in *S. cerevisiae* CEN.PK 113-9D leading to the strain *S. cerevisiae* IBT100082.

The system that was applied for the construction of *S. cerevisiae* IBT100083 was the bipartite gene-targeting technique (Erdeniz et al., 1997). This system is based on a bipartite substrate in which each of the two fragments carries a target sequence, a sequence which needs to be inserted, and a selectable marker gene which is unfunctional but homologous to some part of the same marker in the second fragment. (Figure 3.2)

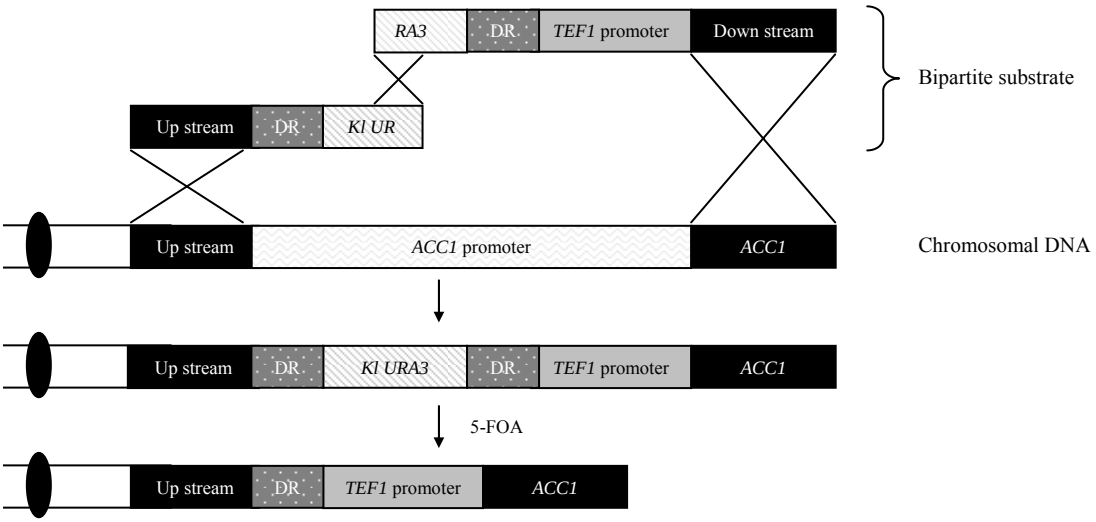


Figure 3.2. The homologous recombination of up and down stream sequences from the bipartite gene-targeting substrate to the chromosomal locus results in the exchange of *ACC1* promoter to *TEF1* promoter and the insertion of *Kl URA3* flanked by direct repeats (DR). The *Kl URA3* was later removed by plating the strains on medium containing 5-fluoroorodic acid (5-FOA).

The first fragment contained the upstream sequence of *ACC1*, direct repeat and the upstream 2/3 *Kluvyromyces lactis* (*Kl*) *URA3*. The upstream which corresponded to the sequence in front of *ACC1* promoter was amplified by primers SWA3 and 4 using genomic DNA of wild type *S. cerevisiae* as a template. Direct repeat with the upstream 2/3 *Kl URA3* was amplified from pWJ1042 as template with primers SWA5 and 6. Two PCR products were then fused together by primers SWA3 and 6.

To obtain the second fragment, three PCR products were generated. The downstream 2/3 *Kl URA3*, which was homologous to 1/3 of *Kl URA3* in the first fragment, was amplified by using primer SWA7 and 8 from pWJ1042. The *TEF1* promoter sequence and the downstream fragment that was homologous to the front part of *ACC1* were amplified from yeast genomic DNA by using primers SWA9/10 and SWA11/12, respectively. The three PCR products were then fused by another PCR reaction using primers SWA7 and 12.

After the fusion, two fragments were transformed into *S. cerevisiae* CEN.PK 113-9D, and integrated into the yeast chromosome by homologous recombination resulting in the changing of the *ACC1* promoter to the *TEF1* promoter. The transformants were plated on the medium containing 5-fluoroorodic acid (5-FOA) in order to pop out the *URA3* so that *URA3* can be used as a marker for the selection again. Since Ura3 metabolises 5-FOA into a toxic compound (Boeke et al., 1984), yeast that maintain *URA3* are killed, whereas yeast lacking *URA3* are resistant to 5-FOA and stay alive, designated as *S. cerevisiae* CEN.PK ACC. Plasmids pRS426CTMSA-PP and pRS424CTnpgA were co-transformed in *S. cerevisiae* CEN.PK ACC to give the strain *S. cerevisiae* IBT100083.

3.3.3 Characterization of IBT100081, 100082 and 100083 in batch cultivations

The physiology of *S. cerevisiae* IBT100081, 100082, and 100083 were characterized in batch cultivations in glucose minimal media with two replicates. Samples were taken during the fermentations for the analysis of biomass dry weight, 6-MSA, and primary metabolites namely ethanol, acetate, succinate, glycerol, and pyruvate in order to determine the kinetics and stoichiometry of the microbial growth. The results from the batch cultivations of the three different strains are shown in Figure 3.3.

Not surprisingly the reference strain had a shorter lag phase than the other strains as it was not burdened by expressing the 6-MSAS and *npgA* genes. The maximum specific growth rate on glucose of the reference strain (IBT100081) was 0.33 hr^{-1} which was higher than for the other two strains, which had specific growth rates of 0.26 and 0.27 hr^{-1} for IBT100082 and IBT100083, respectively.

In the first exponential phase where glucose was being consumed, the production of primary metabolites e.g. ethanol, pyruvate, succinate, glycerol, acetate as well as biomass, CO_2 and 6-MSA were used to calculate yields as C-mmol/C-mol glucose and the results are listed in Table 3.3. The formula $CH_{1.8}O_{0.5}N_{0.2}$ was used for the biomass composition (Roels, 1983). Based on the calculated yields the carbon balance was found to close within 0.4%. Comparison of the three strains showed that there was a positive correlation between the yield of pyruvate, succinate, and glycerol and the yield of 6-MSA, e.g. the yield of pyruvate, succinate and glycerol increased when the yield of 6-MSA on glucose increased. As expected, the yield of 6-MSA was improved when *ACC1* was expressed from the *TEF1* promoter (IBT100083), 3.4 C-mmol/C-mol glucose compared with 2.4 C-mmol 6-MSA/C-mol glucose for the other 6-MSA producing strain (IBT100082). No 6-MSA was produced from the reference strain as it was only carrying the empty plasmids.

Considering only the 6-MSA producing strains, the maximum specific growth rate of the *ACC-TEF1* promoter strain (IBT100083) and *ACC*-native promoter (IBT100082) were about the same (0.27 and 0.26 hr^{-1}). The amount of biomass dry weight formed during the fermentations was therefore approximately the same for both strains, e.g. approximately 145 C-mmol biomass/C-mol glucose. However the yield of 6-MSA and all primary metabolites was higher for IBT100083 than for IBT100082. In contrast, the yield of CO_2 from IBT100083 was significantly lower than that of IBT100082, e.g. 259 and 283 C-mmol CO_2/C-mol glucose, respectively.

After glucose was exhausted, a short lag phase was observed, followed by a second exponential growth phase where the cells were growing on ethanol. During the batch cultivations, samples were also taken from this phase for biomass dry weight, 6-MSA and

primary metabolites (Figure 3.3). The maximum specific growth rate of each strain, as well as the yields of 6-MSA and succinate was also calculated for this growth phase and the results are shown in Table 3.4. In this phase the specific growth rates of the three strains were approximately five times lower than those of the first growth phase. This happens because ethanol is a poorer carbon source than glucose. In contrast, the yield of biomass, succinate and 6-MSA in the ethanol phase were higher than those found in the glucose growth phase. Especially the 6-MSA yield of the IBT100083 strain was much higher (41 C-mmol/C-mol ethanol) compared to the IBT100082 strain (23 C-mmol 6-MSA/C-mol ethanol). Thus the impact of over-expression of the *ACC1* gene is much higher than during growth on glucose, where the yield only increased from 2.4 C-mmol 6-MSA/C-mol glucose to 3.4 C-mmol 6-MSA/C-mol glucose upon over-expression of A*CC1*. Thus, ethanol is clearly a favorable carbon source for the production of 6-MSA. The maximum titer of 6-MSA for IBT100083 was 250 mg/L which was 60% higher than the titer of 150 mg 6-MSA/L obtained with the strain having the native *ACC1*-promoter.

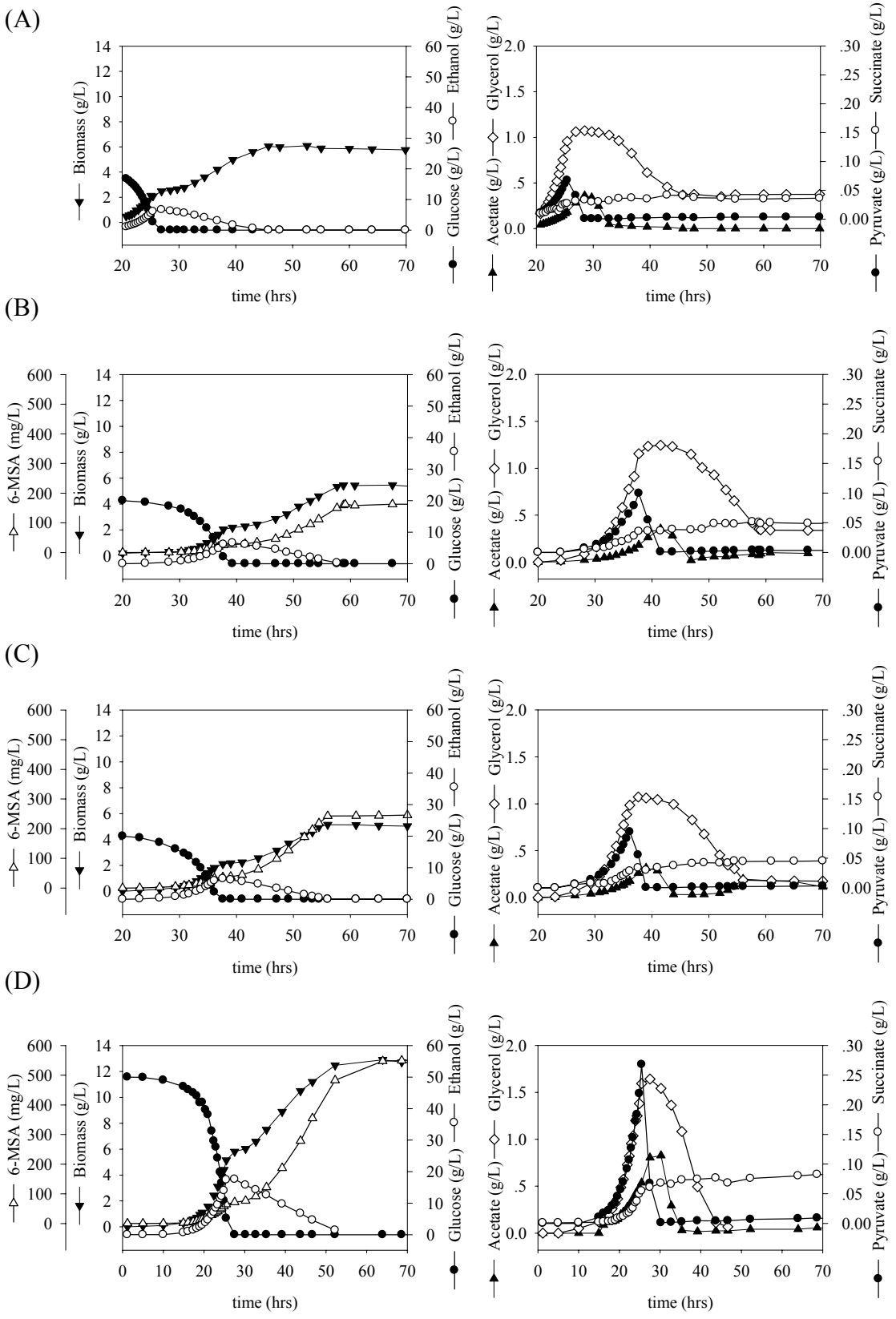

(A)
(B)
(C)
(D)
6-MSA (mg/L)
Biomass (g/L)
Glucose (g/L)
Ethanol (g/L)
Acetate (g/L)
Glycerol (g/L)
Pyruvate (g/L)
Succinate (g/L)
time (hrs)

Figure 3.3. Results from batch cultivations on glucose minimal medium with the three different strains. The left panel shows the concentrations of biomass, glucose, ethanol and 6-MSA (except for the strain with the empty plasmids) and the right panel shows the concentrations of the primary metabolites acetate, glycerol, pyruvate and succinate.

A[1,2]. Results with IBT100081, carrying the empty plasmids pRS424 and pRS426.

B[1,2]. Results with IBT100082, carrying 6-MSAS gene from *P. patulum* and PPTase gene from *A. nidulans*.

C[1,2]. Results with IBT100083, carrying 6-MSAS gene from *P. patulum* and PPTase gene from *A. nidulans*, promoter of *ACC1* was changed to *TEF1* promoter.

D[1,3]. Results with IBT100083, carrying 6-MSAS gene from *P. patulum* and PPTase gene from *A. nidulans*, promoter of *ACC1* was changed to *TEF1* promoter.

[1)] The concentration profiles were very similar for the two cultivations. Data shown is for one of the replicates.

[2)] The glucose concentration was 20 g/L.

[3)] The glucose concentration was 50 g/L.

Table 3.3. Maximum specific growth rate and product yields in the batch cultivations for three different strains during the first exponential phase where glucose was consumed

Maximum specific growth and yields in the first exponential phase	Strains / concentration of glucose			
	IBT100081	IBT100082	IBT100083	
	20 g/L	20 g/L	20 g/L	50 g/L
μ_{max} (hr^{-1})	0.33 ± 0.01	0.26 ± 0.01	0.27 ± 0.00	0.27 ± 0.01
$Y_{S,X}$ (C-mmol/C-mol glc)	151.8 ± 2.8	145.4 ± 0.1	145.7 ± 1.6	142.6 ± 1.4
$Y_{S,EtOH}$ (C-mmol/C-mol glc)	498.6 ± 1.9	495.1 ± 14.5	504.7 ± 2.1	506.7 ± 3.2
$Y_{S,Ace}$ (C-mmol/C-mol glc)	10.2 ± 0.3	9.1 ± 1.1	11.2 ± 0.7	11.3 ± 0.9
$Y_{S,Pyr}$ (C-mmol/C-mol glc)	4.4 ± 0.0	5.6 ± 0.6	6.1 ± 0.2	6.1 ± 0.2
$Y_{S,Gly}$ (C-mmol/C-mol glc)	59.3 ± 1.2	61.2 ± 7.0	63.4 ± 0.0	35.4 ± 0.6
$Y_{S,Suc}$ (C-mmol/C-mol glc)	1.4 ± 0.2	1.8 ± 0.2	1.9 ± 0.1	1.3 ± 0.2
$Y_{S,CO2}$ (C-mmol/C-mol glc)	277.4 ± 6.0	283.4 ± 5.6	259.8 ± 2.7	272.7 ± 3.6
$Y_{S,6\text{-}MSA}$ (C-mmol/C-mol glc)	0.0	2.4 ± 0.1	3.4 ±0.0	2.4 ± 0.1
C-balance (%)	100.3	100.4	99.6	97.9

Table 3.4. Maximum specific growth rate and product yields in the batch cultivations for three different strains during the second exponential phase where ethanol was consumed

Maximum specific growth and yields in the second exponential phase	Strains / concentration of glucose			
	IBT100081	IBT100082	IBT100083	
	20 g/L	20 g/L	20 g/L	50 g/L
μ_{max} (hr^{-1})	0.06 ± 0.00	0.05 ± 0.00	0.06 ± 0.00	0.03 ± 0.00
$Y_{S,X}$ (C-mmol/C-mol EtOH)	535.8 ± 18.3	462.9 ± 34.6	487.0 ± 20.0	464.2 ± 25.6
$Y_{S,Pyr}$ (C-mmol/C-mol EtOH)	0.4 ± 0.0	0.3 ± 0.1	0.3 ± 0.0	0.3 ± 0.0
$Y_{S,Suc}$ (C-mmol/C-mol EtOH)	1.5 ± 0.0	2.3 ± 0.3	2.1 ± 0.1	1.4 ± 0.2
$Y_{S,6\text{-}MSA}$ (C-mmol/C-mol EtOH)	0.0	23.4 ± 5.7	40.9 ± 0.0	36.8 ± 1.5

3.3.4 Theoretical calculation of acetyl-CoA drain for 6-MSA production

Normally the cytosolic acetyl-CoA generated in yeast is mainly used for synthesis of fatty acids that are polymerized to lipids and used in biomass formation. Synthesis of 1 g yeast lipids requires 25.1 mmol of acetyl-CoA (Oura, 1972) and the lipid content is typically between 3 and 12% of the yeast biomass (Verduyn, 1991). However, when 6-MSA production is taking place acetyl-CoA is also needed for 6-MSA biosynthesis. In order to produce 1 mol 6-MSA, in total 4 mol acetyl-CoA is required. To get an idea of how much 6-MSA production would affect the metabolism we calculated the ratio of acetyl-CoA needed for 6-MSA biosynthesis in relation to the need for biomass (Table 3.5). There seems to be a significant drain of acetyl-CoA for 6-MSA biosynthesis since as much as 66% of the amount needed for biomass was used for 6-MSA production in strain IBT100083 when grown on ethanol even if a high lipid content of 10% was assumed for the yeast biomass.

Table 3.5 Theoretical calculation of amount of acetyl-CoA needed for 6-MSA production in relation to amount of acetyl-CoA needed for biomass.

Strain	Phase	Yx,AcCoA (mmol AcCoA for 6-MSA production / g DW)	% of AcCoA to 6-MSA compared to amount that is used for lipid biosynthesis if lipid content is		
			3%	5%	10%
IBT100083	Glucose	0.45	60	36	18
	Ethanol	1.67	222	133	66
IBT100082	Glucose	0.33	44	27	13
	Ethanol	0.85	114	68	34

3.3.5 The effect on 6-MSA production by a higher glucose concentration

As the 6-MSA production was substantially higher with the IBT100083 strain than with any of the other strains, IBT100083 was further characterized by cultivating it in minimal media with 50 g/L of glucose. Samples were taken throughout the batch cultivations for biomass, primary metabolites and 6-MSA analysis (Figure 3.3-D). This data were used to calculate yields on glucose and they are collected in Table 3.3 and compared with the batch cultivation data of the same strain when 20 g/L of glucose was used as a carbon source.

In the first exponential phase, IBT100083 attained the same maximum specific growth rate of about 0.27 hr^{-1} when it was growing in the media containing 20 and 50 g/L glucose. Furthermore, the yields on glucose for most of the primary metabolites such as pyruvate, ethanol and acetate, as well as biomass, were pretty much the same in both fermentations. The yields of succinate and glycerol were significantly higher in the cultivations with 20 g/L than with 50 g/L of glucose, but the yield of CO_2 showed the opposite result as it was higher in the 50 g/L glucose cultivation. However, the yield of 6-MSA dropped from 3.35 to 2.41 C-mmol/C-mol glucose when the initial glucose concentration was increased from 20 and 50 g/L.

In the second exponential growth phase where the cells consumed ethanol, the maximum specific growth rate was approximately 50% lower (0.03 compared with 0.06 hr^{-1}) in the

cultivation with the initial glucose concentration of 20 compared with that of 50 g/L of glucose. Moreover, a decrease in the biomass, succinate, and 6-MSA yields were observed when 50 g/L of glucose was applied. However, still a high final titer of 6-MSA was obtained, namely 600 mg/L.

3.4 Discussion

Due to its GRAS status, well-established molecular biology and fermentation techniques, *S. cerevisiae* is a popular model organism for eukaryotic cell biology, but it is also a widely used cell factory for the production of proteins, and various kinds of metabolites ranging from the primary metabolites such as ethanol and succinic acid to secondary metabolites like 6-MSA (Keyley et al., 1998; Wattanachaisaereekul et al., 2007), lycopene (Yamano et al., 1994), and the antimalarial drug precursor or artemisinic acid (Ro et al., 2006). Several recombinants of *S. cerevisiae* have been successfully used for the over-expression of these products. This can be achieved by regulating the expression or co-expression of the genes using replicating expression vectors that are maintained at high copy numbers in the cells and using strong promoters that are either inducible or constitutive to allow the expression of the genes at a high level.

The *GAL1* promoter which is considered as a very strong inducible promoter has been extensively used for expression of heterologous genes in *S. cerevisiae*. It was also used in our previous study for the heterologous expression of 6-MSAS from *P. patulum* and PPTase from *Bacillus subtilis* and *A. nidulans* in *S. cerevisiae* (Wattanachaisaereekul et al., 2007). The *GAL1* promoter is tightly repressed by glucose but it is strongly induced by galactose. Therefore, in order to express the genes, galactose, which is an expensive carbon source, must be added. Even more, the expression of the strong *GAL1* promoter may lead to toxic effects (Mumberg et al., 1994) as it has been observed that a steady state could not be obtained in chemostat cultivations while *GAL1* promoter was used for expression of 6-MSAS and PPTase (Chapter 4). All these reasons made us construct a new vector system that involved a glucose-based promoter.

For this purpose the *TEF1* promoter, the promoter of a translation elongation factor gene was used. This is a strong constitutive promoter, and the advantage of using the *TEF1* promoter over the *GAL1* promoter is that the heterologous gene could be expressed during growth on glucose which is a cheap carbon source, and one of the most abundant sugars in nature. Moreover, the specific growth rate of *S. cerevisiae* is much higher during growth on glucose compared with growth on galactose. Comparison of using the *TEF1* promoter (IBT100082) in this study and using the *GAL1* promoter to express 6-MSA and PPTase gene in our previous study (Wattanachaisaereekul et al., 2007), illustrates that although the yield of 6-MSA in the first exponential phase is higher with galactose as carbon source (8.6 C-mmol/C-mol galactose) than with glucose as carbon source (2.4 C-mmol/C-mol glucose), the glucose promoter results in the same final titer of about 150 mg/L, but as the specific growth rate is faster this final titer is reached faster, e.g. the productivity is higher with glucose as carbon source.

To further increase the production of 6-MSA we aimed for increasing the supply of the precursors for 6-MSA biosynthesis. The precursors of 6-MSA biosynthesis are acetyl-CoA, malonyl-CoA, and NADPH. Therefore, in order to have a strain producing more 6-MSA, these precursors must be increased by for example performing the genetic modification that can redirect the fluxes towards these precursors. In *S. cerevisiae*, the acetyl-CoA pool is prolific, whereas sources of malonyl-CoA are generally limited. For that reason, the *ACC1* gene which encodes the enzyme acetyl coenzyme A carboxylase was chosen in this study to be overexpressed, as it catalyzes the conversion of acetyl-CoA to malonyl-CoA. Over-expression of *ACC1* by changing its promoter to the strong constitutive *TEF1* promoter enhanced the biosynthesis of malonyl-CoA, leading to the high level production of 6-MSA as shown in Table 3.3 and 3.4. In other words, by changing the promoter of *ACC1* to *TEF1* promoter in *S. cerevisiae* carrying plasmids with the 6-MSAS gene and *npgA* (IBT100083), the production of 6-MSA was increased by about 60%. The positive effect of overexpressing *ACC1* was more pronounced during growth on ethanol compared to growth on glucose. Snf1 is known to be activated during growth in glucose-derepressing conditions. Snf1 deactivates Acc1 by phosphorylation (Woods et al 1994; Shirra et al., 2001) and hence more of Acc1 is likely to be inactivated in ethanol growth

conditions. Overexpression of acetyl-CoA carboxylase in *Streptomyces coelicolor* has also been seen to result in an improved polyketides production (Ryu et al., 2006).

IBT100083 was further studied in batch cultivations containing 50 g/L of glucose, and the results revealed that by using 50 g/L of glucose, the yeast was stressed and therefore the yield of 6-MSA decreased, but still the overall production of 6-MSA was good as a final titer of 600 mg/L could be reached. Thus, we conclude that it is possible to produce high levels of fungal polyketides in *S. cerevisiae*, and this organism hence represent an interesting platform for future production of novel fungal polyketides.

3.5 References

Boeke JD, LaCroute F, Fink GR. 1984. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol Gen Genet. 197(2): 345-346.

Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. 1992. Multifunctional yeast high-copy-number shuttle vectors. Gene 110: 119-122.

Dawson MJ, Farthing JE, Marshall PS, Middleton RF, O'Neill MJ, Shuttleworth A, Stylli C, Tait M, Taylor PM, Wildman HG, Buss AD, Langley D, Hayes MV. 1992. The squalestatins, novel inhibitors of squalene synthase produced by a species of *Phoma*. J. Antibiotics 45: 639-647.

Erdeniz N, Mortensen UH, Rothstein R. 1997. Cloning-free PCR-based allele replacement methods. Genome Res. 7: 1174-1183.

Gietz RD, Woods RA. 2002. Transformation of yeast by the LiAc/SS carrier DNA/PEG method. Methods in Enzymology 350: 87-96.

Katz L, Donadio S. 1993. Polyketide synthesis: Prospects for hybrid antibiotics. Annu. Rev. Microbiol. 47: 875-912.

Kealey JT, Liu L, Santi DV, Betlach MC, Barr PJ. 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proc. Natl. Acad. Sci. 95: 505-509.

Mumberg D, Müller R, Funk M. 1994. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Research 22(25): 5767-5768.

O'Hagen D. 1991. The polyketide metabolites. Ellis Harwood, London.

O'Neill JA, Simpson TJ, Willis CL. 1993. Biosynthesis of colletodiol and related polyketide macrodiolides in *Cytospora* sp. ATCC 20502: Synthesis and metabolism of advanced intermediates. J. Chem. Soc., Chem. Commun. 738-740.

Oura E. 1972. The effect of aeration on the growth energetics and biochemical composition of baker's yeast. Ph.D. thesis, University of Helsinki, Helsinki.

Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MCY, Withers ST, Shiba Y, Sarpong R, Keasling JD. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440(7086): 940-943.

Roels JA. 1983. Energetics and kinetics in biotechnology. Amsterdam: Elsevier.

Ryu YG, Butler MJ, Chater KF, Lee KJ. 2006. Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*. Appl Environ Microbiol. 72(11): 7132-7139.

Sambrook J, Russell DW. 2001. Molecular Cloning: A laboratory manual (3rd editon). Cold Spring Harbor Laboratory Press, NY, USA.

Shirra MK, Patton-Vogt J, Ulrich A, Liuta-Tehlivets O, Kohlwein SD, Henry SA, Arndt KM. 2001. Inhibition of acetyl coenzyme A carboxylase activity restores expression of the *INO1* gene in a *snf1* mutant strain of *Saccharomyces cerevisiae*. Mol Cell Biol. 21(17): 5710-5722.

Turner WB, Aldridge DC. 1970. Metabolites of *Helminthosporium monoceras*: structures of monocerin and related benzopyrans. J. Chem. Soc. 18: 2598-2600.

Verduyn C. 1991. Physiology of yeasts in relation to biomass yields. Antonie van Leewenhoek 60: 325-353.

Wattanachaisaereekul S, Lantz AE, Nielsen ML, Andrésson ÓS, Nielsen J. 2007. Optimization of heterologous Production of the polyketide 6-MSA in *Saccharomyces cerevisiae*. Biotechnol Bioeng. 97(4): 893-900.

Woods A, Munday MR, Scott J, Yang X, Carlson M, Carling D. 1994. Yeast SNF1 is functionally related to mammalian AMP-activated protein kinase and regulates acetyl-CoA carboxylase in vivo. J. Biol. Chem. 269: 19509-19516.

Yamano S, Ishii T, Nakagawa M, Ikenaga H, Misawa N. 1994. Metabolic engineering for production of beta-carotene and lycopene in *Saccharomyces cerevisiae*. Biosci. Biotechnol. Biochem. 58(6): 1112-1114.

Chapter 4

Towards Stable Production of 6-MSA by *S. cerevisiae* in Chemostat Cultivation

---

Abstract

Saccharomyces cerevisiae has become increasingly interesting for being employed as a cell factory for the heterologous production of several different metabolites, e.g. polyketides. Usually foreign genes encoding the enzymes required for biosynthesis can be expressed in yeast by the introduction of shuttle vectors or plasmids. In continuous cultivations, however, plasmid instability leading to the loss of the plasmids from the cells and a decline in product formation may occur. Therefore, construction of stable plasmids and strains which maintain stable expression of the genes during the cultivation is of interest. In this study, the stability of a plasmid-based strain producing the polyketide 6-MSA was examined. It was found that continuous cultivations with *S. cerevisiae* IBT100075 carrying two plasmids containing genes encoding 6-MSAS and PPTase, respectively, with both genes under the control of the *GAL1* promoter, did not allow for stable production of 6-MSA. Even at low dilution rates the production of 6-MSA dramatically decreased after a maximum production that only lasted a few generations. In order to construct a stable production strain, the 6-MSAS gene was integrated into the chromosome with the co-expression of *npgA*, encoding PPTase from *A. nidulans*, and using the *TEF1* promoter for controlling the expression of the genes. This resulted in a stable production strain which could be cultivated at high dilution rates on a glucose minimal medium with a stable production of 6-MSA for several generations.

4.1 Introduction

In commercial production of heterologous proteins using *S. cerevisiae* as production host, plasmid based expression of the biosynthetic genes is often used, and a key problem associated with this is plasmid instability. Plasmid stability is defined as the ability of transformed cells to maintain the plasmid unchanged during their growth, manifesting their phenotypic characteristics (Imanaka and Aiba, 1981), and therefore, instability is the tendency of the transformed cells to lose their engineered properties because of changes to or loss of plasmids. This problem is especially significant for recombinant yeasts as almost all yeast plasmid vectors are hybrid plasmids or shuttle vectors that are relatively unstable. In addition, the instability can be caused by uneven partitioning of plasmids during cell division (Primrose and Ehrlich, 1981). The stability of a plasmid is also a function of other genetics factors such as the plasmid copy number, the presence of selective markers, the strength of promoter used in the plasmid, and environmental factors like medium composition, temperature and dilution rate (DiBiasio and Sardonini, 1986). The combination of these factors may result in instability of the plasmid, and this can be an important factor in the use of a recombinant strain for large scale production.

As mentioned before, almost all plasmid vectors for *S. cerevisiae* are shuttle vectors. In fact they are both yeast and *E. coli* vectors that can replicate in either of the two organisms. This kind of vectors usually consists of all or part of an *E. coli* vector and a yeast replication system. Yeast autonomous plasmid vectors are based on one of two replication systems, either the sequence derived from the endogenous yeast 2 μm plasmid or a chromosomal DNA fragments that can replicate autonomously. (Kingsman et al., 1979; Stinchcomb et al., 1979). The later type is usually the autonomously replicating sequence (ARSs) which presumably acts as origins of replication in yeast chromosomes. Several types of yeast vectors have been reported by Panchal (1987). The principal types and properties of yeast vectors are summarized in Table 4.1.

Table 4.1. Type and properties of vectors used in *S. cerevisiae*[1]

Vector	Transformation efficiency/ μg of DNA	Copy number per cell	Mitotic stability	Elements of yeast DNA present
Yeast Integrating plasmid (YIp)	1-10	1	Very stable	Selectable marker, chromosomal fragment
Yeast Episomal plasmid (YEp)	10^3 - 10^5	5-40	Slightly unstable	Selectable marker, 2 μm origin
Yeast Replicating plasmid (YRp)	10^3 - 10^5	3-30	Very unstable	Selectable marker, chromosomal origin (ARS)
Yeast Centromeric plasmid (YCp)	10^3 – 10^4	1	Stable	Selectable marker, chromosomal or 2 μm origin, centromere (CEN)
Yeast Linear plasmid (YLp)	10^3 – 10^4	5-30	Very unstable	Selectable marker, chromosomal or 2 μm origin, telomere (TEL)

[1]Modified from Saunders et al. (1986)

The most widely used plasmid and the best compromise in terms of stability and transformation efficiency appears to be offered by YEps. However, these plasmids are slightly unstable when the recombinant yeast carrying this kind of plasmid undergoes many rounds of duplication. This leads to problems for large scale production situations where it is necessary to propagate to high biomass concentrations. An alternative plasmid for introduction of DNA into yeast is the integrating plasmids (YIps). Even though the DNA integrated into the host genome by recombination shows low transformation efficiency the result is a high stability. In this case, however, the copy number of the introduced DNA is only one, so the expression levels are generally lower than for high copy number plasmids.

In this study we investigate the stability of heterologous production of the polyketide 6-MSA in *S. cerevisiae* during growth in continuous cultures The strain IBT100075, which is carrying the YEp plasmids harbouring the 6-MSAS gene from *P. patulum* and the *sfp* gene encoding the enzyme phosphopantetheinyl transferase (PPTase) from *B. subtilis*, both under control of the strong inducible *GAL1* promoter, was earlier constructed and characterized in batch cultures (Wattanachaisaereekul et al., 2007). In chemostat

cultivation, however, this strain turned out to be unstable, possibly due to the reasons mentioned above, and furthermore the use of multiple copies of the *GAL1* promoter resulted in very slow growth of the strain, and hence it could only be operated at very low dilution rates. We therefore constructed another strain (IBT100084) by using an integrating vector for the insertion of the 6-MSAS gene into the yeast chromosome at the YMLCΔ2 site. However, the IBT100084 strain was still carrying the gene encoding PPTase (*npgA* from *A. nidulans)* on a plasmid. In IBT100084 both the 6-MSAS and the PPTase encoding genes are under control of the *TEF1* promoter, allowing the strain to grow and produce 6-MSA in glucose minimal media and at high dilution rates. The production of 6-MSA from IBT100084 was investigated as well as the stability of the strain was assessed and compared to the strain IBT100075 in order to build a yeast based cell factory platform for the production of polyketides.

4.2 Materials and methods

4.2.1 Molecular biology techniques

Oligonucleotides used were synthesized by MWG-Biotech AZ, Germany, and are listed in Table 4.2. PCR reactions were performed using Phusion polymerase and Taq polymerase according to the manufacturer's guideline. GFX spin columns from Amersham Pharmacia were used for purification of DNA from either gel bands or solution. Total genomic DNA extraction from *S. cerevisiae* for PCR template, DNA ligation, restriction endonuclease digestion and analysis were done according to the standard protocols.

Transformation of yeast cells was carried out according to Gietz and Woods (2002). *E. coli* DH5α was used for propagation of recombinant plasmids. The preparation of competent *E. coli* and transformation protocol were performed according to Sambrook and Russell (2001).

Table 4.2. List of primers used in this study. The underline sections indicate the restriction sites introduced to facilitate cloning.

DNA fragment	primers	Sequence 5′-3′
		*Cla*I
CYC1 terminator	SWA13	CCATCGATTCATGTAATTAGTTATGTCACGCT
		*Kpn*I
	SWA14	GGGGTACCGGCCGCAAATTAAAGCCTTCGAGC
		*Sac*I
Direct repeat of pRS306	SWA15	CGAGCTCCAGTCGGGAAACCTGTCGTGCCAG
		*Not*I *Fse*I *Asi*SI
	SWA16	AAAGCGGCCGCGGCCGGCCGCGATCGCTCGTGCACACAGCCCAGCTTGGAG
		*Fse*I
Upstream of *YMLCΔ2*	SWA17	ATCGCGAGTCTTCATTCACCAGTCGGCCGGCCGGTGATCCTTCCGATGTATGCGTG
		*Not*I
	SWA18	AAAGCGGCCGCATGGATTCCTAAATCCTCGAGGAG
		*Asi*SI
Downstream of *YMLCΔ2*	SWA19	GCGATCGCCAATCCTTGCGTTTCAGCTTCCAC
		*Fse*I
	SWA20	GGCCGGCCGACTGGTGAATGAAGACTCGCGATTGCATCACAAGCACCCGGTACTGC
		*Not*I
TEF1 promoter	SWA21	AAGGAAAAAAGCGGCCGCCACACACCATAGCTTCAAAATGTT
		*Xba*I
	SWA22	GCTCTAGATTTGTAATTAAAACTTAGATTAGA

4.2.2 Plasmids and strains

The plasmids and stains used in this study are listed in Table 4.3.

Table 4.3. List of plasmids and strains used in this study

	Characteristics	Reference
Plasmids		
pRS306	Yeast integrating shuttle vector with *URA3* marker	Sikorski and Hieter (1989)
pRS424CTnpgA	Expression vector containing *TEF1* promoter, *CYC1* terminator, *TRP1* marker, and *npgA* from *A. nidulans*	Wattanachaisaereekul et al. (submitted)
pGEM-7Zf(+)/LIC-F	Vector permitting visual detection of recombinants	Haun et al. (1992)
pGei425HSA-PP	Carrying 6-MSAS gene from *P. patulum* with *GAL1* promoter, *CYC1* terminator, and *LEU2* marker	Wattanachaisaereekul et al. (2007)
pRS306CRUDTMSA-PP	Integrating vector carrying *CYC1* terminator, direct repeats, Up-Down sequence of integration site – *YMLCΔ2*, *TEF1* promoter, *URA3* marker and 6-MSAS gene from *P. patulum*	This study

Strains		
CEN.PK113-9D	MATa;*ura3-52;HIS3;LEU2;trp1-289;MAL2-8;SUC2*	Peter Kötter [a]
IBT100075	CEN.PK 111-27B with plasmids pGei425HSA-PP and p424sfp	Wattanachaisaereekul et al. (2007)
IBT100084	CEN.PK 11-39 with 6-MSAS integrated at YMLCΔ2 and pRS424CTnpgA	This study

[a] Institut für Mikrobiologie, der Johann Wolfgang Goethe-Universität Frankfurt, Germany

4.2.3 Construction of an integrating vector expressing 6-MSAS

A plasmid for integration into the yeast chromosome was constructed (Figure 4.1). The *CYC1* terminator was amplified from the genomic DNA of the wild type *S. cerevisiae* CEN.PK 113-7D with primers SWA13 and 14. The PCR product was digested with *Cla*I and *Kpn*I, and ligated into pRS306 in the *Cla*I/*Kpn*I sites, resulting in the plasmid pRS306C. The direct repeat was amplified using pRS306 as a template with primers SWA15 and 16 by Taq polymerase. In order to facilitate the ligation in the next step, the PCR product was first ligated into pGEM-T easy vector systemI (Promega Corporation, Madison, WI, USA) and digested by *Sac*I and *Not*I. The released *Sac*I-*Not*I fragment was then ligated into *Sac*I/*Not*I of the plasmid pRS306C to yield pRS306CR. Introduction of the direct repeat enables looping out of the *URA3* marker after integration so that the same vector and marker may be used repeatedly

The upstream and downstream, which are homologous to the upstream and downstream of the *YMLCΔ2* integration site, were amplified by the primer pairs SWA17/18 and SWA19/20, respectively. The fragments were fused together using primer SWA18 and 19, so that the starting point of the upstream fragment was followed by the end of the downstream fragment with a restriction site, *Fse*I, in between. After the fusion, the fragment was subcloned into pGEM-T easy vector systemI and cut by *Not*I and *Asi*SI. The *Not*I-*Asi*SI fragment was introduced into pRS306CR to obtain pRS306CRUD. To construct plasmid pRS306CRUDT, the *TEF1* promoter was amplified from yeast genomic DNA by primers SWA21 and 22. The PCR product was cut by *Not*I and *Xba*I and ligated into the plasmid pRS306CRUD.

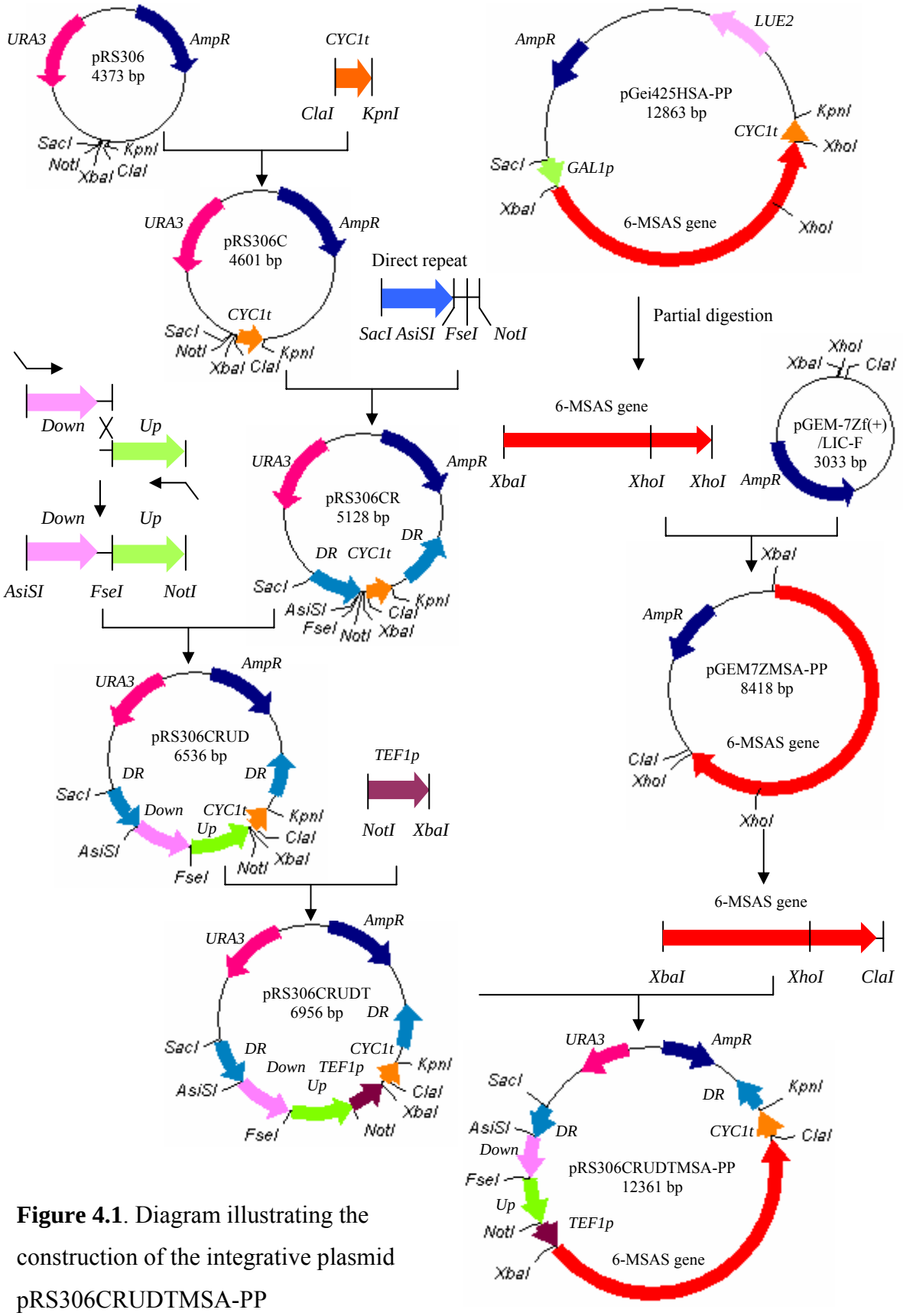


Figure 4.1. Diagram illustrating the construction of the integrative plasmid pRS306CRUDTMSA-PP

The gene encoding 6-MSAS from *P. patulum* was excised from p425GeiHSA-PP by completed digestion with *Xba*I and partial digestion with *Xho*I, as a *Xho*I restriction site also is present inside the gene. The fragment with the 6-MSAS gene was subcloned into the *Xba*I/*Xho*I site of plasmid pGEM-7Zf(+)/LIC-F in order to change the restriction sites to enable ligation into the plasmid pRS306CRUDT. The 6-MSA gene was cut out by *Xba*I and *Cla*I and introduced into pRS306CRUDT to yield the integrative plasmid pRS306CRUDTMSA-PP. Restriction enzyme digestions verified the map of the constructed vector.

4.2.4 Bioreactor and cultivation conditions

The strains were cultivated in 100 mL medium in 500 mL baffled shake flasks for propagation. The medium consisted of glucose or galactose 10 g/L depending on the experiment, $(NH_4)_2SO_4$ 7.5 g/L, KH_2PO_4 14.4 g/L, $MgSO_4.7H_2O$ 0.5 g/L, 0.05 mL/L Antifoam 298 (Sigma-Aldrich, St. Louis, MO, USA) and 2 mL/L trace metal solution ($FeSO_4.7H_2O$ 3 g/L, $ZnSO_4.7H_2O$ 4.5 g/L, $CaCl_2.6H_2O$ 4.5 g/L, $MnCl_2.2H_2O$ 0.84 g/L, $CoCl_2.6H_2O$ 0.3 g/L, $CuSO_4.5H_2O$ 0.3 g/L, $Na_2MoO_4.2H_2O$ 0.4 g/L, H_3BO_3 1 g/L, KI 0.1 g/L, and $Na_2EDTA.2H_2O$ 15 g/L), and 2 mL/L vitamin solution (d-biotin 50 mg/L, Ca-pantothenate 1 g/L, thiamin-HCl 1 g/L, pyridoxin-HCl 1 g/L, nicotinic acid 1 g/L, p-aminobenzoic acid 0.2 g/L, and m-inositol 12.5 g/L). pH of the medium was adjusted to 6.5 by 2M NaOH prior to autoclaving. The cultures were incubated at 30 ^oC, with shaking at 150 rpm (model 3033, GFL, Burgwedel, Germany). When the absorbance (A_{600}) reached 13 (2.3 g biomass/L), 0.1 mL of the culture was used as inoculum for the initial batch cultivations.

The initial batch cultivations were carried out in bioreactors (Biostat Braun Biotech International GmbH, Mulsungen, Germany) with a working volume of 1 L. When the concentration of CO_2 measured in the off-gas by a gas analyzer (Innova AirTech Instruments A/S, Ballerup, Denmark) after peaking decreased to 0.8%, the cultures were switched from the initial batch phase to chemostat mode with a dilution rate of 0.03 hr^{-1} for chemostat no.1, 2 and 3, and 0.1 hr^{-1} for chemostat no. 4. The cultures were fed with

minimal medium containing glucose or galactose 20 g/L, $(NH_4)_2SO_4$ 10 g/L, KH_2PO_4 3 g/L, $MgSO_4.7H_2O$ 1 g/L with 0.05 mL/L Antifoam 298, 2 mL/L of trace metal solution and 2 mL/L of vitamin solution. The compositions of the trace metal and vitamin stock solutions were the same as mentioned above. The temperature of the cultivations was kept constant for the whole cultivation at 30 ^oC, and the pH was automatically controlled at 5.0 by automatic addition of 2M KOH. Aerobic conditions were maintained by sparging the cultures with air, 1 L/min, and by having an agitation of 500 rpm. After starting the feed, the steady states were determined by constant biomass dry weight, absorbance of the cultures and concentration of CO_2.

4.2.5 Absorbance and biomass measurement

Absorbance was measured at 600 nm in a Hitachi U-1100 spectrophotometer (Tokyo, Japan) for biomass determination of the inoculum before initial batch cultivations. Biomass dry weight was determined throughout the cultivations by filtering a known volume of fermentation broth through a dried, pre-weighed nitrocellulose filter (Gelman Science, Ann Arbor, MI, USA) with a pore size 0.45 μm. The residue was washed twice with distilled water. The filter was dried to constant weight in a microwave oven at 150 W for 10 minutes, cooled down in a desiccator, and the weight gain was measured.

4.2.6 Analysis of glucose, galactose and 6-MSA

Fermentation broth was filtered through a 0.45 μm pore-sized cellulose acetate filter (GE Osmonics Labstore, Minnetonka, MN, USA) for later analysis of substrate and products. Glucose and galactose were analyzed by high-performance liquid chromatography (Dionex-HPLC; Sunnyvale, CA), equipped with an Aminex HPX-87H ion exclusion column, 300 mm × 7.8 mm (Bio-Rad Laboratories, Hercules, CA), which was operated at 60 ^oC, and a flow rate of 0.6 mL/min of 5 mM H_2SO_4, using a refractive index detector (Shodex RI-71, Tokyo, Japan). 6-MSA was quantified from the filtrate by HPLC (Agilent 1100 series) equipped with a Luna C18(2) column (150 × 4.60 mm, 5 micron particles), using a gradient of 50 ppm trifluoroacetic acid (TFA) in milliQ water (solvent A) and 50

ppm TFA in acetonitrile (solvent B) at a flow rate of 1 mL/min. The gradient of the solvents was 20% to 60% B in 10 minutes, then 20% B for the next 2 minutes.

4.3 Results

4.3.1 Characterization of the plasmid-based strain producing 6-MSA in chemostat cultivation

The cultivations of *S. cerevisiae* IBT100075 was initiated with glucose minimal medium in order to prove that 6-MSA production was repressed by glucose (Figure 4.2). Another reason was to assess the stability of the strain when no 6-MSA was produced from the cell. In the end of the batch phase, upon glucose exhaustion, feeding of a minimal medium containing 20 g/L of glucose was started to inititate the chemostat cultivation. The strain IBT100075 have previously been characterized in batch cultivations (Wattanachaisaereekul et al., 2007), and the maximum specific growth rate (μ_{max}) of this strain in a galactose minimal medium was 0.1 hr^{-1}. The dilution rate for the chemostat cultivation was therefore chosen to be quite low, 0.03 hr^{-1}. For a short period, from 150 hr to 200 hr, constant conditions were achieved when glucose was the carbon source (Figure 4.2). However, the biomass concentration then started to decrease even though no 6-MSA was produced. After 340 hours, corresponding to approximately 10 generations, in the glucose minimal medium, the feed was switched to a galactose minimal medium containing 20 g/L of galactose (Figure 4.2). The dilution rate was kept at 0.03 hr^{-1}. As expected, 6-MSA started to be produced just after the switch to galactose with the maximum titer being 110 mg/L between 400-430 hours. Shortly after this both biomass and 6-MSA concentrations dramatically dropped, indicating that no steady state could be obtained from this cultivation.

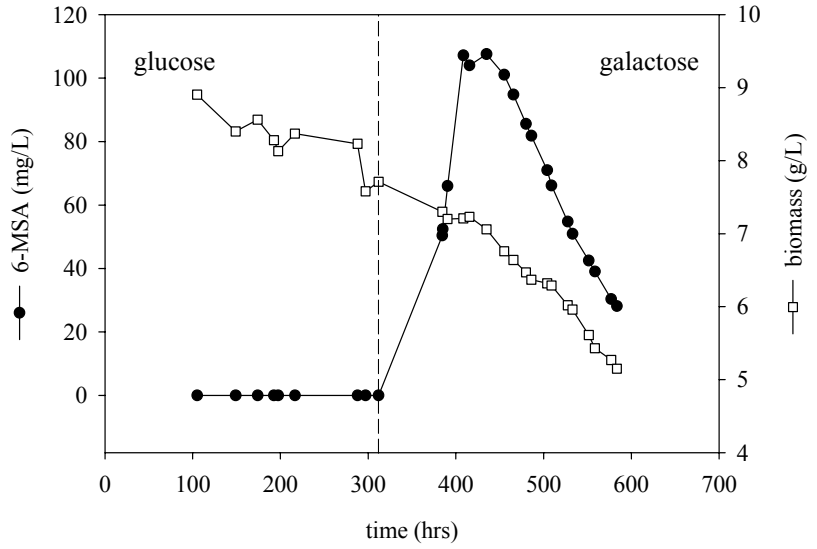


Figure 4.2. The amount of biomass dry weight and 6-MSA produced by *S. cerevisiae* IBT100075 in chemostat cultivation (no.1) using glucose and galactose minimal media (carbon source 20 g/L) with a dilution rate of 0.03 hr^{-1}.

The next chemostat cultivation was performed using the same strain and it was also initiated with a glucose minimal medium (glucose 20 g/L) and the glucose phase was kept for longer to see whether a steady state could be achieved when no production was taking place (Figure 4.3). After 1250 hours, the feed was changed to a galactose minimal medium (galactose 20 g/L). The dilution rate was constant 0.03 hr^{-1} throughout the cultivation. The cell mass dry weight concentration profile shows that a steady state could be obtained on the glucose minimal medium but not on the galactose minimal medium. Yet, even for growth on glucose the pattern indicated occurrence of instability phenomena. 6-MSA was produced after the feed was changed to the media containing galactose and the maximum level of 6-MSA was 150 mg/L at 1400 hours. This production level was, however, only kept for a very short period of time, thereafter the production of 6-MSA decreased to zero and the cellmass concentration increased, but then started to decrease again.

In order to check if the plasmids were lost from the cells, and hence contributing to the observed instability, samples taken from the fermentation were plated on both YPD and a synthetic medium without leucine and tryptophan, and the total number of yeast colony

forming unit were counted on both plates. Cells that have lost the plasmids can grow on YPD medium, which contains all amino acids, but not on a synthetic medium without leucine and tryptophan, which are required for the survival of the cells. It seemed as if the plasmids were lost during the fermentation since the number of yeast colony forming units on the synthetic media was much lower than the number on YPD medium plates. The total number of cells per mL in the sample harvested from the chemostat cultivation (no. 2) at 1570h that could grow on YPD and a synthetic medium, respectively, were about 1.70×10^8 and 8×10^6 colonies/ml.

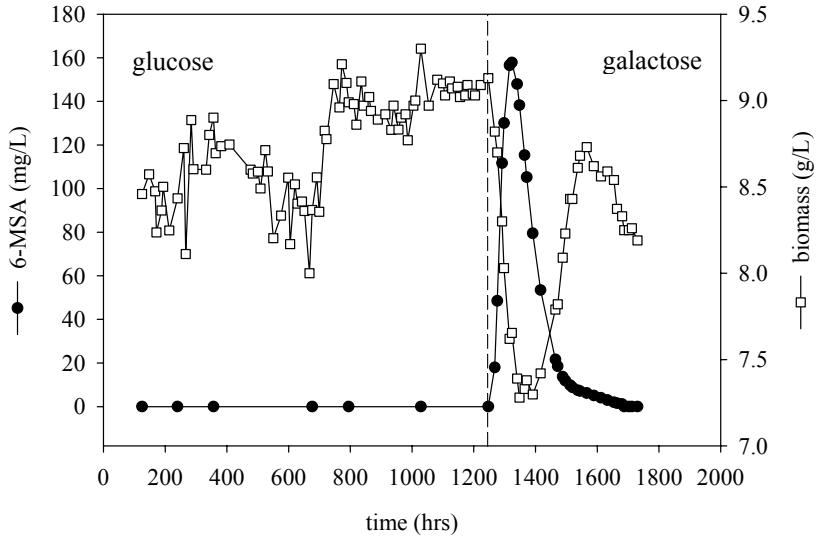


Figure 4.3. The amount of biomass dry weight and 6-MSA produced by *S. cerevisiae* IBT100075 in chemostat cultivation (no.2) using glucose and galactose minimal media (carbon source 20 g/L) with the dilution rate of 0.03 hr^{-1}.

The last 1L chemostat cultivation performed with the plasmid-based strain (IBT100075) was started directly with a galactose minimal medium (galactose 20 g/L), still with a dilution rate of 0.03 hr^{-1}. The results of the biomass dry weight and production of 6-MSA from this cultivation are shown in Figure 4.4, which shows that the strain produced 6-MSA and that a steady state of 6-MSA production could be obtained, for a short period of time, approximately 150 hours corresponding to 4.5 generations. At this period, the strains yielded 125.5±1.6 mg of 6-MSA/L, after that the concentration of 6-MSA decreased until

the end of the cultivation. To our surprise, during the period of stable 6-MSA production, the cellmass concentration was not stable, but had an increasing trend.

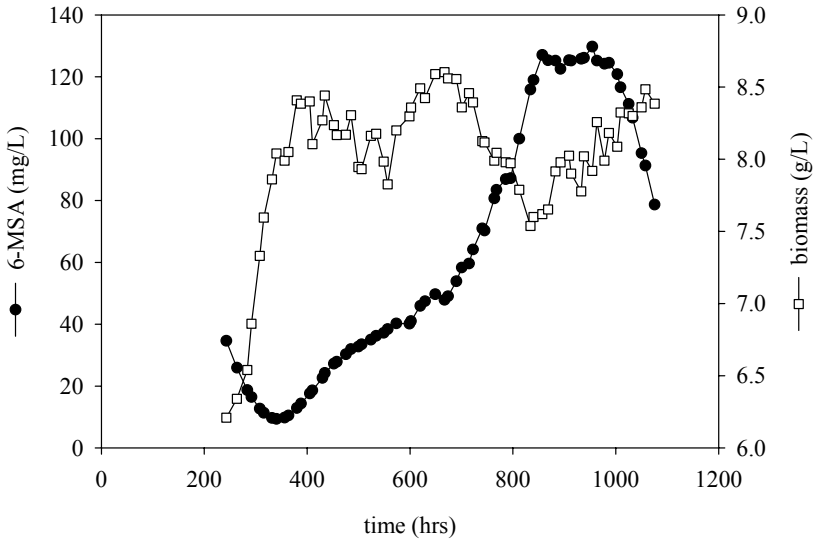


Figure 4.4. The amount of biomass dry weight and 6-MSA produced by *S. cerevisiae* IBT100075 in chemostat cultivation (no.3) using galactose minimal media (20 g/L) with the dilution rate of 0.03 hr^{-1}.

4.3.2 Construction of a recombinant strain expressing 6-MSAS

Since *S. cerevisiae* IBT100075 carrying YEp plasmids with the genes encoding 6-MSAS from *P. patulum* and PPTase from *B. subtilis*, both under the control of the *GAL1* promoter, was unstable in chemostat cultivation, the gene encoding 6-MSAS was integrated into the yeast chromosome. An integrative vector pRS306CRUDTMSA-PP, carrying the 6-MSAS gene from *Penicillium patulum* under expression of the *TEF* promoter and *URA3* as selectable marker was constructed (Figure 4.1). To construct the strain IBT100084, the plasmid pRS306CRUDTMSA-PP was linearized with *Fse*I and transformed into *S. cerevisiae* CEN.PK 113-9D, resulting in the integration of the 6-MSAS gene in the chromosome by homologous recombination. The integration of the 6-MSAS gene was checked by colony PCR using the primer SWA22 and the primer homologous to the sequence outside the upstream region. The size of PCR was correct

(data not shown) indicating that the plasmid was indeed integrated in the right site. The strain was subsequently transformed by plasmid pRS424CTnpgA, a multi-copy vector carrying a gene encoding phosphopantetheinyl transferase and transformants were selected on medium without uracil and tryptophan.

4.3.3 Characterization of the 6-MSA integrated strain

In order to verify that the effect of the integration itself, an integrating vector without the 6-MSAS gene was also integrated at the YMLCΔ2 site of *S. cerevisiae* CEN.PK 113-5D and this strain was used as reference. Batch cultivation with a glucose minimal medium (glucose 20 g/L) was performed in order to check whether the integration has an effect on the growth and compare to the growth of the wild type strain (CEN.PK 113-7D). The results showed that the maximum specific growth rate (μ_{max}) of the integrated strain without the 6-MSAS gene was 0.34±0.020 hr^{-1}, whereas that of wild type was 0.37±0.004 hr^{-1}. This means that the integrated plasmid, and thereby the deletion of the YMLCΔ2 site, did not seem to cause any severe phenotype on the strain.

S. cerevisiae IBT100084 was able to produce 6-MSA during growth in a glucose minimal medium as the integrated 6-MSAS gene and the gene encoding PPTase on the plasmid were under control of the strong constitutive *TEF1* promoter. The strain was first characterized in a batch cultivation with a glucose minimal medium (glucose 20 g/L) and this resulted in maximum production of 6-MSA of 16.0±0.8 mg/L with a μ_{max} of 0.31±0.011 hr^{-1} (data not shown), which is considerably higher than the growth rate 0.1 hr^{-1} obtained on galactose for the plasmid strain, hence enabling chemostat cultivations with higher dilution rates. Chemostat cultivation in a glucose minimal medium (20 g/L) was then carried out with a working volume of 1L and a dilution rate of 0.1 hr^{-1}. The result from this chemostat cultivation shown in Figure 4.5 demonstrates that it was possible to obtain steady-state. The cultivation lasted for approximately 20 generations from 50-250 hours with a production of 6-MSA resulting in a titer of 14.3±0.8 mg/L. One might argue that there was a slight increase in the values both for cellmass and 6-MSA, and hence that two different steady-states were seen over the time period (Figure 4.5). The first steady

state would then be from 50 -100 hours with a cellmass concentration of 5.7±0.06 g/L and the second one 170-245 hours with a cellmass concentration of 6.4±0.05 g/L.

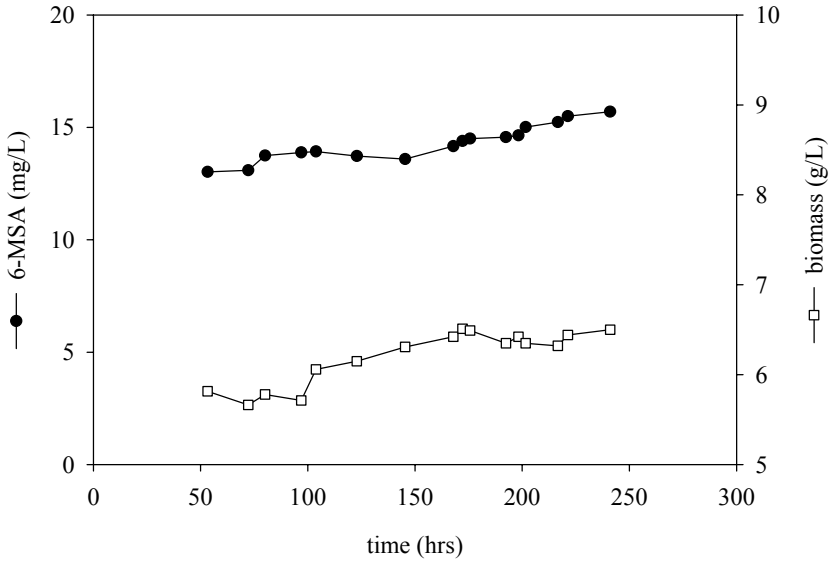


Figure 4.5. The amount of biomass dry weight and 6-MSA produced by *S. cerevisiae* IBT100084 (6-MSAS integrated strain) in chemostat cultivation (no.4) using glucose minimal media (glucose 20 g/L) with the dilution rate of 0.1 hr^{-1}.

4.4 Discussion

Plasmid stability is clearly desirable for application of recombinant strains in the production of new products. Here we examined the stability of *S. cerevisiae* producing the polyketide 6-MSA. We found that *S. cerevisiae* IBT100075 carrying two YEp plasmids with the 6-MSAS gene from *P. patulum* and the PPTase gene from *B. subtilis*, did not allow stable production in a chemostat cultivation with neither glucose nor galactose minimal media, even though the plasmids used in this strain contained the endogenous yeast 2 μm, which should ensure a high copy number and is generally considered to be the best replication of origin for maintaining stability. However, there are several factors involved in causing instability of this strain, such as the copy number of the plasmid, the partition system, the expression level or transcription efficiency effect and the dilution rate.

From the literature it is known that the stability of plasmids in *S. cerevisiae* is affected by the plasmid copy number. Because the desired gene is usually inserted into the plasmid, the copy number of the plasmid determines the dosage of the gene in the cell. In general, plasmids showing high copy number have been shown to exhibit greater stability than those with low copy number (Futcher and Cox, 1984). During cell division, efficient partitioning of plasmids between mother and daughter cells will prevent emergence of plasmid-free cells. On the other hand, inefficient partitioning will result in plasmid-free cells. Based on this concept, Walmsley et al. (1983) reasoned that a plasmid with a high copy number should generate plasmid-free cells with a lower frequency than when the plasmid copy number is low. However, there do not seem to be a direct effect of copy number on stability as it has been found that the stability of a YRp plasmid (YRp7M) in a haploid strain was significantly lower than that of a YEp plasmid (pMA3a), although the former contained 952±391 copies of YRp7M per haploid genome, whereas the latter which was more stable contained only contains 133±59 copies of pMA3a (Spalding and Tuite, 1989). Thus, a high copy number does not necessarily guarantee the stability of the plasmid, but it seems to be more important with an effective partitioning system. Therefore, the instability of the strain harboring the plasmids with the 6-MSAS gene and the PPTase gene in the chemostat cultivations performed here might partly be due to a defect in the partitioning of the plasmid DNA between daughter and mother cells during cell division, which will result in a gradual reduction in the copy number, and eventually a loss of the plasmid from the cells and hereby a decline in the production of 6-MSA.

The effect of dilution rate on plasmid stability in chemostat cultivation has been extensively studied. Since the dilution rate corresponds to the specific growth rate in the continuous culture, the chemostat cultivation provides an easy method for elucidating the relationship between specific growth rate and plasmid stability. Several studies showed that plasmid stability increased with increasing dilution rate when *S. cerevisiae* containing a plasmid was continuously cultivated in a selective media (DiBiasio and Sardonini, 1986; Lee and Hassan, 1987; Parker and DiBiasio, 1987; Lee and Hassan, 1988; Caunt et al., 1989). One reason being that the partitioning of the plasmid DNA between daughter and mother cells during cell division breaks down during slow, glucose-limited growth

(Bugeja et al., 1989). Furthermore, at low growth rates normally a higher copy-number and hence increase in plasmid expression is observed. This indicates that replication and transcription of the plasmid have the potential to damage the plasmid DNA (Parker and DiBiasio, 1987). In this case, plasmid stability would depend on the cell's DNA repair capabilities. Because many of the enzymes required for DNA repair are synthesized in different periods of time, the repair capability or stability of plasmid is likely to vary with the stage of growth cycle. Also the higher protein burden resulting from the higher expression level probably contributed to the instablility. Plasmid-strains with high expression of recombinant enzymes have earlier been shown to be more unstable that strains with low expression (Meinander and Hahn-Hägerdal, 1997). In accordance, the large size of the 6-MSA expressing plasmid could be another instablility issue. It has earlier been demonstrated that larger plasmids are considerably more instable (Futcher and Cox, 1984; Gerbaud et al., 1981). Moreover, it has been reported that the expression of the strong *GAL1* promoter could lead a toxic effects in the cells, probably due to an out-titration of *GAL4* (Mumberg et al., 1994). This could also contribute to the instability of plasmids in the cells observed in this study.

Due to the difficulty to maintain the steady state of the plasmid-harboring strain producing 6-MSA in chemostat cultivation, we constructed a new strain (IBT100084). Instead of using YEp plasmids, we used a very stable integrative YIp plasmid (pRS306) for the integration of the 6-MSAS gene into the chromosome. We used YMLCΔ2 as the integration site since it is located in a well-expressed area, close to the *GAL80* gene on chromosome XIII. It is also a non-functional open reading frame, and no function will therefore be lost when the gene encoding 6-MSAS is integrated at this site. Moreover, there are many more of this type integration sites in chromosome, so if more than one copy of the 6-MSAS gene is needed, it can be integrated into other sites by changing the homologous sequence according to the sequence of the site. However, the new strain (IBT100084) was still carrying PPTase on a plasmid. The gene encoding for PPTase in the plasmid had been changed from *sfp* in the previous strain (IBT100075) to *npgA* in this strain (IBT100084) as we reported before that the expression of PPTase from *A. nidulans* (NpgA) leaded to substantially higher production of polyketide than Sfp from *B. subtilis*

(Wattanchaisaereekul et. al., 2007). In addition, we also changed the promoter of the 6-MSAS gene and the gene encoding for PPTase from the *GAL1* promoter to the strong constitutive *TEF1* promoter in the new strain. This was done in order to avoid the potential toxic effects to the cells from the use of *GAL1* promoter and to be able to cultivate the strain in a glucose minimal medium and hence apply higher dilution rates.

After cultivating the strain IBT100084 in the chemostat with a dilution rate of 0.1 hr^{-1}, we found that it was possible to obtain a steady state. When comparing the production of 6-MSA between the integrated strain and the plasmid containing strain, the integrated strain yielded 14.3 mg of 6-MSA/L or 0.23 mg 6-MSA/g dw/hr, whereas the plasmid strain yielded 125.5 mg of 6-MSA/L or 0.47 mg of 6-MSA/g dw/hr. Even though the titer was much lower with the integrated strain, the productivity of 6-MSA was only two-fold lower. One should also take into consideration that the integrated strain only contains a single copy of the 6-MSAS gene in the genome, whilst there are more copies of the plasmid in the plasmid strain. This indicates the advantage of the integrated strain cultivating in a glucose minimal medium at a high dilution rate and possibly also the use of the PPTase from *A. nidulans*.

In summary, we obtained a stable *S. cerevisiae* strain, which could produce 6-MSA in chemostat cultivations. The problem with the instability of the plasmids from the plasmid-based strain was overcome by the integration of the 6-MSAS gene into the chromosome Both the 6-MSAS gene and the PPTase gene were under control of the strong constitutive *TEF1* promoter instead of using *GAL1* promoter in order to be able to produce 6-MSA in glucose-based medium and with a high dilution rate. This strain is suitable to apply in industrially relevant conditions, e.g. long time fed-batch processses to obtain high biomass concentrations and hence high 6-MSA titers with glucose based media. The 6-MSA titer obtained with the integrated strain was much lower than that with the plasmid containing strain, but as the strain could be operated at a higher dilution rate and the productivity was only slightly lower and it may probably be improved through integration of additional copies of the polyketide synthetase encoding gene into the chromosome.

4.5 References

Bugeja VC, Kleinman MJ, Stanbury PF, Gingold EB. 1989. The segregation of the 2 mu-based yeast plasmid pJDB248 breaks down under conditions of slow, glucose-limited growth. J Gen Microbiol. 135(11): 2891-2897.

Caunt P, Impoolsup A, Greenfield PF. 1989. The effect of oxygen limitation and stability of a recombinant plasmid in *Saccharomyces cerevisiae*. Biotechnol. Lett. 11: 5-10.

DiBiasio D, Sardonini C. 1986. Stability of continuous culture with recombinant organism. Ann NY Acad Sci 469: 111-117.

Futcher AB, Cox BS. 1984. Copy number and the stability of 2 μm circle-based artificial plasmids of *S. cerevisiae*. J. Bacteriol 157: 283-290.

Gerbaud C, Elmerich C, Tandeau de Marsac N, Chocat P, Charpin N, Guerineau M, Aubert JP. 1981. Construction of new yeast vectors and cloning of the Nif (nitrogen fixation) gene cluster of *Klebsiella pneumoniae* in yeast. Curr. Genet. 3: 173-180.

Gietz RD, Woods RA. 2002. Transformation of yeast by the LiAc/SS carrier DNA/PEG method. Methods in Enzymology 350: 87-96.

Haun RS, Serventi IM, Moss J. 1992. Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. BioTechniques 13: 515-518.

Imanaka T, Aiba S. 1981. A perspective on the application of genetic engineering: stability of recombinant plasmid. Ann. NY. Acad. Sci. 369: 1-14.

Kingsman AJ, Clarke L, Mortimer RK, Carbon J. 1979. Replication in *S. cerevisiae* of plasmid pBR313 carrying DNA from the yeast TRP1 region. Gene 7: 141-152.

Lee FJS, Hassan HM. 1987. Effect of oxygen tension on stability and expression of a killer toxin chimeric plasmid in a chemostat culture of *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 27: 72-74.

Lee FJS, Hassan HM. 1988. Stability and expression of a plasmid-containing killer toxin cDNA in batch and chemostat cultures of *Saccharomyces cerevisiae*. Biotechnol. Bioeng. 31: 783-789.

Meinander NQ, Hahn-Hägerdal B. 1997. Fed-batch xylitol production with two recombinant *Saccharomyces cerevisiae* strains expressing XYL1 at different levels, using glucose as a cosubstrate : A comparison of production parameters and strain stability. Biotechnol. Bioeng. 54(4): 391-399.

Mumberg D, Müller R. Funk M. 1994. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Research 22(25): 5767-5768.

Panchal CJ. 1987. Yeast plasmids and transformation. *In* Berry DR. (ed.) Yeast Biotechnology. Allen&Unwin, London, pp 80-103.

Parker C, DiBiasio O. 1987. Effect of growth rate and expression level on plasmid stability in *Saccharomyces cerevisiae*. Biotechnol. Bioeng. 29: 215-221.

Primrose SB, Ehrlich SD. 1981. Isolation of plasmid deletion mutants and study of their instability. Plasmid 6: 193-201.

Sambrook J, Russell DW. 2001. Molecular Cloning: A laboratory manual (3rd editon). Cold Spring Harbor Laboratory Press, NY, USA.

Saunders G, Tuite MF, Holt G. 1986. Fungal cloning vectors. Trends Biotechnol 4: 93-98.

Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics 122: 19-27.

Spalding A, Tuite M. 1989. Host-plasmid interactions in *Saccharomyces cerevisiae*: effect of host ploidy on plasmid stability and copy number. J. Gen. Microbiol. 135: 1037-1045.

Stinchcomb DT, Struhl K, Davis RW. 1979. Isolation and characterization of a yeast chromosomal replicatior. Nature 282: 39-43.

Wattanachaisaereekul S, Lantz AE, Nilesen ML, Andrésson OS, Nielsen J. 2007. Optimization of heterologous production of the polyketide 6-MSA in *Saccharomyces cerevisia.* Biotechnol Bioeng. 97(4): 893-900.

Wattanachaisaereekul S, Lantz AE, Nielsen ML, Nielsen J. 2007. Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. (Submitted).

Walmsley RM, Gardner DC, Oliver SG. 1983. Stability of a cloned gene in yeast grown in chemostat culture. Mol. Gen. Genet. 192: 361-365.

Chapter 5

Effect of Glucose Concentration on the Production of the Polyketide 6-MSA by *S. cerevisiae* in Batch Cultivations

---

Abstract

In a previous study it was shown possible to obtain high titers of the heterologously produced 6-MSA using the yeast *S. cerevisiae* as cell factory. This was obtained during growth on glucose, which is the preferred carbon source for industrial fermentations. In order to evaluate whether the final titer could be further enhanced we therefore optimized the production of 6-MSA by cultivating the *S. cerevisiae* IBT100083 in glucose minimal media with different initial glucose concentration; 20, 50, 100, 150 and 200 g/L. In these fermentations there will be formed different levels of ethanol during the initial fermentative growth phase. Ethanol was earlier shown to be a good carbon source for the polyketide 6-MSA, but high ethanol concentrations will also be toxic for the cells. We found that at an initial glucose concentration between 100-200 g/L, there was poor production of 6-MSA in both the glucose and the ethanol growth phase, possibly due to the hyperosmotic pressure imposed by the high glucose concentration as well as toxicity from ethanol and other byproducts that are formed during the cultivation. In contrast, the yeast presented the high specific growth rate and yielded better 6-MSA in the media with lower concentration of initial glucose, and optimum was found to be around 20-50 g/L.

5.1 Introduction

In the fermentation industry, the yeast *S. cerevisiae* is widely used not only for baking and brewing, but for the production of various primary metabolites such as ethanol and glycerol, heterologous proteins and being used as a host for the production of secondary metabolites, e.g. polyketides. The growth of *S. cerevisiae* depends very much on the substrate especially the carbon source like glucose. Since it is the primary fuel used by most living cells to generate the energy that is needed to be carried out cellular function, glucose plays an essential role in the regulation of the genes and metabolism inside the cell (Fiechter et al., 1981). Like most eukaryotes, *S. cerevisiae* prefers glucose over other sugars, and it rapidly converts glucose into ethanol and CO_2 at both anaerobic and aerobic conditions (Käppeli, 1986; Pronk et al., 1996). At anaerobic conditions, acetaldehyde is the final electron acceptor and fermentation is the only way to generate Gibbs free energy and hereby sustain growth. At aerobic conditions, oxygen may serve as the final electron acceptor through the respiratory pathway, but *S. cerevisiae* still exhibits alcoholic fermentation due to glucose repression of respiration. At these conditions, fermentation and respiration co-exist and the metabolism is referred to as oxido-reductive or respiro-fermentative (Käppeli, 1986).

For heterologous production of the polyketide 6-MSA by *S. cerevisiae*, the cells need not only to express the polyketide synthase gene, but they need to be cultivated aerobically in order to obtain sufficient energy for the biomass synthesis and for formation of the precursors such as acetyl-CoA, malonyl-CoA and NADPH for 6-MSA biosynthesis. We previously reported that the use of C2 carbon source such as ethanol by *S. cerevisiae* resulted in the higher production of acetyl-CoA and NADPH, leading to the substantially higher production of 6-MSA in the ethanol phase, as the maximum yield of 6-MSA from the strain IBT100083 was 34 mg 6-MSA/g ethanol, whereas the yield of 6-MSA on glucose was only 2 mg/g glucose (Chapter 3). This indicates that the occurrence of alcoholic fermentation in aerobic conditions, even though it is definitely undesirable in the biomass-directed applications, provides to a certain extent benefits to the production of polyketides.

The appearance of ethanol in the aerobic conditions is dependent on the initial concentration of glucose. Therefore the objective of this study was to investigate the production of the polyketide 6-MSA at different glucose concentrations in the media, resulting in different levels of production of ethanol which can then be used for polyketide production as previously described. In this study we used 6-MSA as a model polyketide, produced by *S. cerevisiae* IBT100083 with the co-expression of *npgA* from *A. nidulans* and the promoter of *ACC1* of this strain was changed to *TEF1* promoter in order to ensure an efficient supply of malonoyl-CoA. Physiological characterizations of the 6-MSA producing strain was performed during the batch cultivations and the yields of 6-MSA on each cultivation were evaluated.

5.2 Materials and methods

5.2.1 Strains

The strain used throughout this study was *S. cerevisiae* IBT100083. The strain is carrying the two plasmids with the 6-MSAS gene from *P. patulum* and the *npgA* encoding PPTase from *A. nidulans*, and with the *ACC1* promoter being changed to the *TEF1* promoter as previously described in Chapter 3.

5.2.2 Bioreactors and cultivation conditions

The strain was cultivated in 100 mL medium in 500 mL baffled shake flasks for propagation. The medium consisted of glucose 10 g/L, $(NH_4)_2SO_4$ 7.5 g/L, KH_2PO_4 14.4 g/L, $MgSO_4.7H_2O$ 0.5 g/L, 0.05 mL/L Antifoam 298 (Sigma-Aldrich, St. Louis, MO, USA) and 2 mL/L trace metal solution ($FeSO_4.7H_2O$ 3 g/L, $ZnSO_4.7H_2O$ 4.5 g/L, $CaCl_2.6H_2O$ 4.5 g/L, $MnCl_2.2H_2O$ 0.84 g/L, $CoCl_2.6H_2O$ 0.3 g/L, $CuSO_4.5H_2O$ 0.3 g/L, $Na_2MoO_4.2H_2O$ 0.4 g/L, H_3BO_3 1 g/L, KI 0.1 g/L, and $Na_2EDTA.2H_2O$ 15 g/L), and 2 mL/L vitamin solution (d-biotin 50 mg/L, Ca-pantothenate 1 g/L, thiamin-HCl 1 g/L, pyridoxin-HCl 1 g/L, nicotinic acid 1 g/L, p-aminobenzoic acid 0.2 g/L, and m-inositol 12.5 g/L). pH of the medium was adjusted to 6.5 by 2M NaOH prior to autoclaving. The cultures were incubated at 30 ^{o}C, with shaking at 150 rpm (model 3033, GFL, Burgwedel,

Germany). When the absorbance (A_{600}) reached 13 (2.3 g biomass/L), 0.1 mL of the culture was used as inoculum for the batch cultivations.

Batch cultivations were carried out in bioreactors (Biostat Braun Biotech International GmbH, Mulsungen, Germany) with a working volume of 2 L. The medium consisted of glucose 20, 50, 100, 150 and 200 g/L for each cultivation, $(NH_4)_2SO_4$ 10 g/L, KH_2PO_4 3 g/L, $MgSO_4.7H_2O$ 1 g/L with 0.05 mL/L Antifoam 298, 2 mL/L of trace metal solution and 2 mL/L of vitamin solution. The compositions of the trace metal and vitamin stock solutions were the same as mentioned above. The temperature of the cultivations was kept at 30 ^oC, and the pH was automatically controlled at 5.0 by addition of 2M KOH. The agitation was 500 rpm and the bioreactors were aerated with 2 L air/min.

5.2.3 Biomass measurement

Biomass dry weight was determined by filtering a known volume of fermentation broth, approximately 10-15 mL depending on the biomass concentration in the media, through a dried, pre-weighed nitrocellulose filter (Gelman Science, Ann Arbor, MI) with a pore size 0.45 μm. The residue was washed twice with distilled water. The filter was dried to constant weight in a microwave oven at 150 W for 10 minutes, cooled down in a desiccator, and the weight gain was measured.

5.2.4 Analysis of glucose, primary metabolites and 6-MSA

Fermentation broth was filtered through a 0.45 μm pore-sized cellulose acetate filter (GE Osmonics Labstore, Minnetonka, MN) for later analysis of substrate and products. Glucose and primary metabolites were analyzed by high-performance liquid chromatography (Dionex-HPLC; Sunnyvale, CA), equipped with an Aminex HPX-87H ion exclusion column, 300 mm × 7.8 mm (Bio-Rad Laboratories, Hercules, CA), which was operated at 60 ^oC, and a flow rate of 0.6 mL/min of 5 mM H_2SO_4 using a refractive index detector (Shodex RI-71, Tokyo, Japan) and UV detector (UVD340S, Dionex, Softron GmbH, Germering, Germany). 6-MSA was quantified from the filtrate by HPLC (Agilent 1100 series) with Luna C18(2) column, 150 × 4.60 mm with 5 micron of the

porous silica particles to which the C18 phase is bonded, using a gradient of 50 ppm trifluoroacetic acid (TFA) in milliQ water (solvent A) and 50 ppm TFA in acetonitrile (solvent B) at a flow rate of 1 mL/min. The gradient of the solvents was 20% to 60% B in 10 minutes, then 20% B for the next 2 minutes.

5.3 Results

5.3.1 Growth rate, production of 6-MSA and primary metabolites in the first exponential phase

S. cerevisiae IBT100083 was cultivated in batch cultivations with the glucose concentration of 100, 150 and 200 g/L. The results were compared with the batch cultivations with 20 and 50 g/L of glucose that were described in Chapter 3. The growth characteristics and the production of biomass, primary metabolites and 6-MSA as a function of time are shown in Figure 5.1.

During the first exponential phase, the cells consumed glucose at their maximum rate. Even though the same strain was applied throughout this study, it grew at different specific growth rates due to the different initial glucose concentration in the medium. The maximum specific growth rate decreased in the fermentations with high concentrations of glucose, from 0.27 hr^{-1} at 20 g/L to 0.22 hr^{-1} at 200 g/L of glucose. In contrast the specific glucose uptake rate increased from 2.29 g glucose/g dw/hr at 20 g/L to 2.63 g glucose/g dw/hr at 150 g/L of glucose, but it was lower at the highest glucose concentration due to the lower specific growth rate at this glucose concentration (Figure 5.2).

The total biomass and CO_2 formation produced during the first exponential growth phase increased according to the increase in the initial glucose concentration, but the yields on glucose decreased at for increasing glucose concentrations. A similar observation was made for 6-MSA and all the primary metabolites except for pyruvate. In another words, although the total 6-MSA production increased from 6.34 mg/L at 20 g/L of glucose to approximately 100 mg of 6-MSA/L at 200 g/L of glucose, the yields of 6-MSA decreased

from 2.1 to 0.6 mg 6-MSA/g glucose when the glucose concentration increased from 20 to 200 g/L. Unlike for the other primary metabolites, the yield of pyruvate on glucose increased when the concentration of glucose in the media increased until 100 g/L, but it then decreased for higher initial glucose concentrations (Figure 5.2). The increase in the pyruvate yield when the glucose concentration increased to a certain level seems to correlate with the specific glucose uptake rate. This indicates that these two parameters were related to each others. However, all the product yields were calculated in C-mol/C-mol glucose (see Table 5.1) in order to evaluate the carbon balance, and this shows that for high initial glucose concentrations it was difficult to close the carbon balance. This is probably due to the formation of other byproducts such as organic acids and aldehydes during the fermentations with the high initial glucose concentration, and these compounds were not measured by the HPLC.

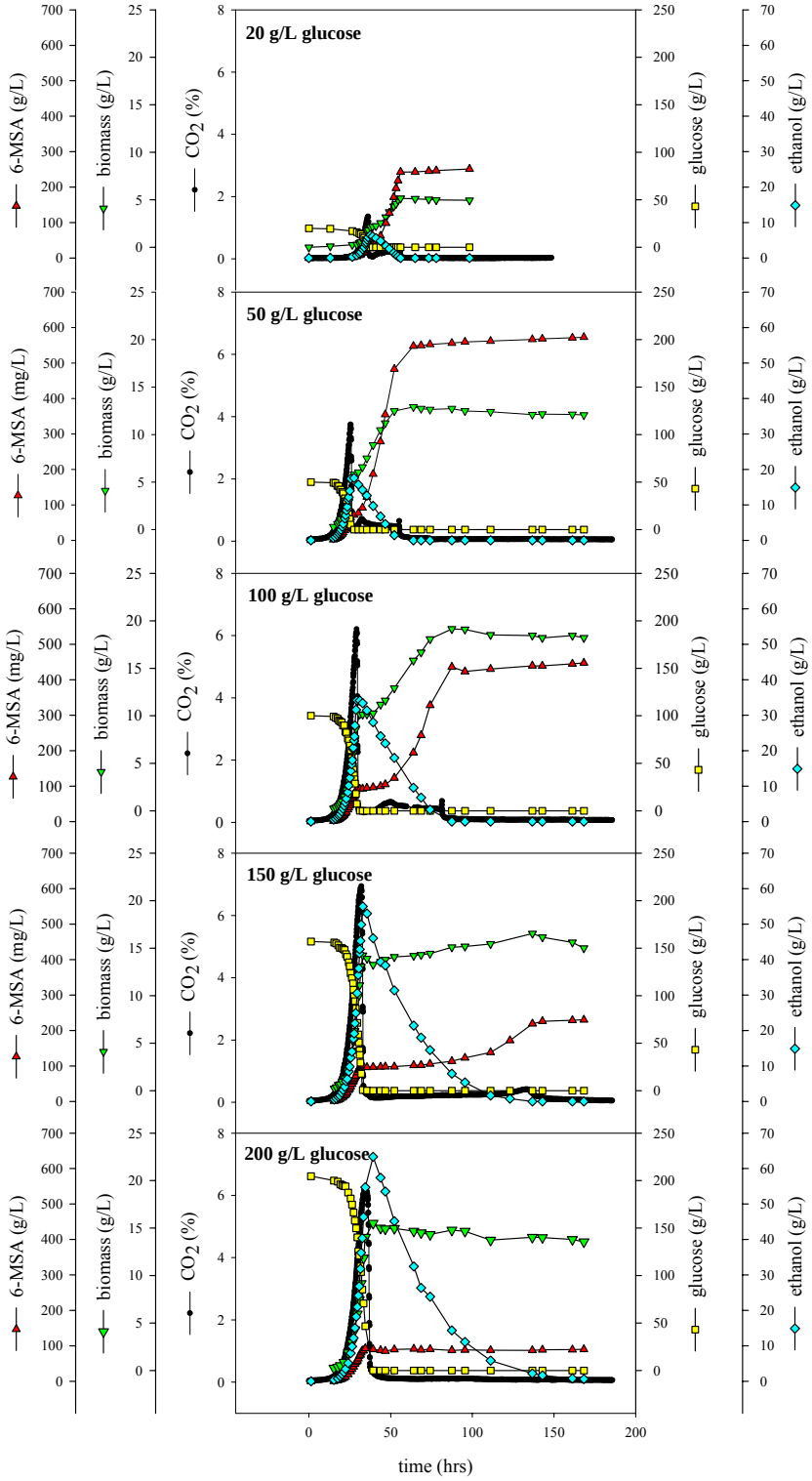


Figure 5.1. The concentration of 6-MSA, biomass, CO_2, glucose and ethanol during the batch cultivations of *S. cerevisiae* IBT100083 in minimal media with different initial glucose concentrations.

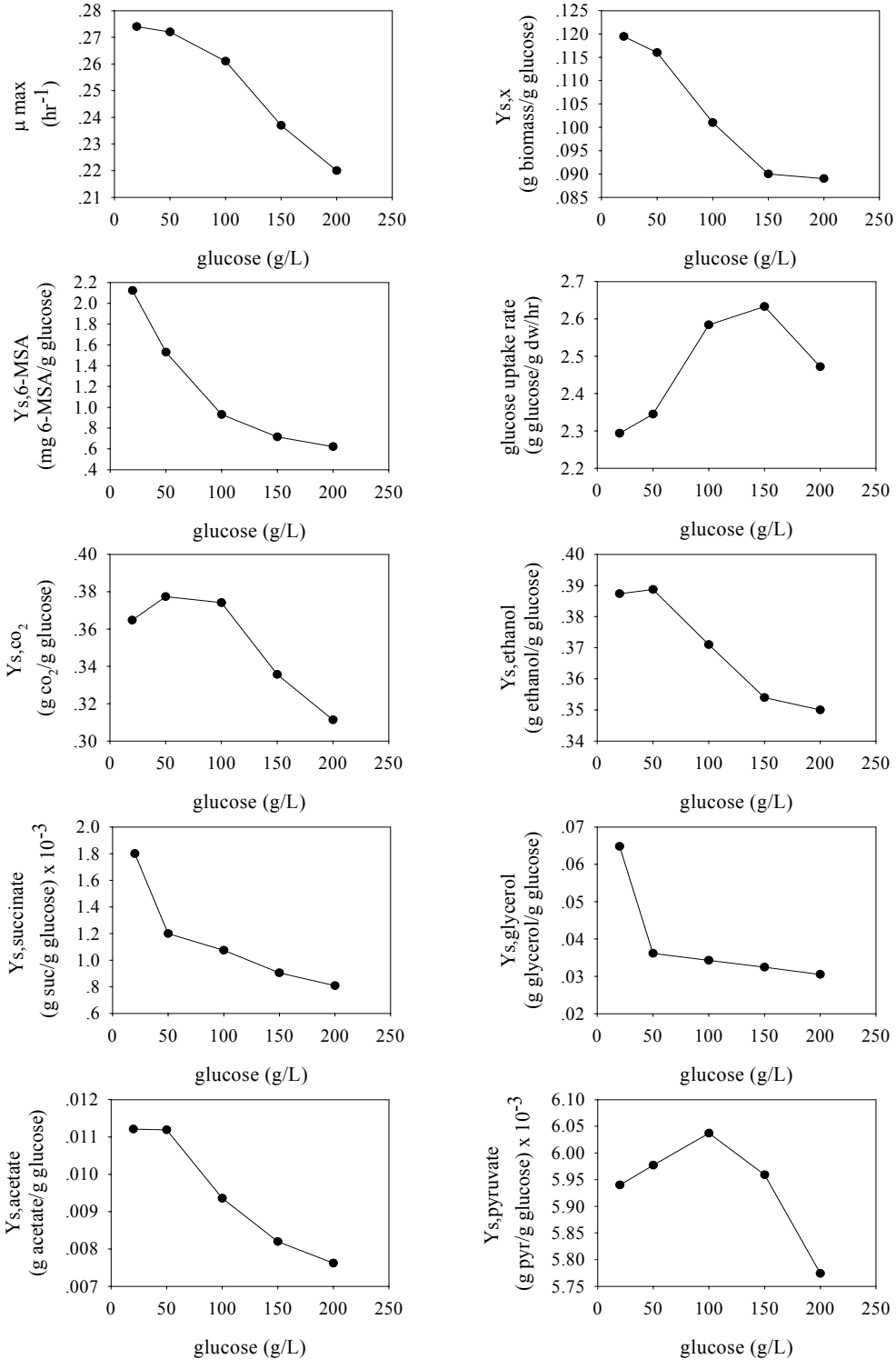


Figure 5.2. The changes in the specific growth rate, specific glucose uptake rate and product yields in the glucose growth phase when growing *S. cerevisiae* IBT100083 in minimal media with different initial glucose concentrations.

Table 5.1. Product yields in glucose phase during the batch cultivations of *S. cerevisiae* IBT100083 in minimal media containing different glucose concentration.

Yields	Initial glucose concentration (g/L)				
(C-mmol/C-mol glucose)	20	50	100	150	200
$Y_{glucose,6\text{-}MSA}$	3.35	2.41	1.47	1.13	0.98
$Y_{glucose,biomass}$	145.68	142.64	123.15	109.74	108.52
$Y_{glucose,ethanol}$	504.96	506.71	483.61	461.45	456.24
$Y_{glucose,succinate}$	1.86	1.28	1.09	0.92	0.82
$Y_{glucose,acetate}$	11.21	11.19	9.36	8.20	7.62
$Y_{glucose,glycerol}$	63.37	35.38	33.55	31.74	29.86
$Y_{glucose,pyruvate}$	6.08	6.11	6.18	6.10	5.91
$Y_{glucose,CO2}$	259.78	272.70	255.27	229.01	212.41
C-balance (%)	99.63	97.84	91.37	84.83	82.24

5.3.2 Growth rate, production of 6-MSA and primary metabolites in the second exponential phase

After glucose was consumed, the cells were growing on ethanol formed from the overflow metabolism in the glucose growth phase. The amount of ethanol in each batch cultivation was different owing to the different initial glucose concentrations. The total amount of ethanol produced from the cultivations after the complete exhaustion of glucose with initial glucose concentration of 20, 50, 100, 150, and 200 g/L were about 6.4, 17, 34, 52, and 63 g ethanol/L, respectively.

As reported before, the use of ethanol as carbon source allowed at least ten-fold higher 6-MSA yield than the use of glucose. However, this seemed to be dependent on the concentration of ethanol in the media. For low concentrations of ethanol, such as 6.4 and 17 g/L, which resulted in the fermentations with initial glucose concentrations of 20 and 50 g/L, the cells grew well and effectively used the ethanol as carbon source, resulting in high 6-MSA final titers of 250 and 570 mg of 6-MSA/L, respectively. The final titer of 6-MSA after the second exponential growth phase started to decrease when the concentration of ethanol in the media was 34 g/L (450 mg 6-MSA/L), and even lower

titers of 230 and 100 mg 6-MSA/L when the ethanol concentration was 52 and 63 g/L, respectively (Figure 5.1).

In the batch cultivation containing 52 g/L of ethanol (150 g/L of initial glucose), the cells grew with a poor specific growth rate as the biomass only increased slowly and also the CO_2 production rate was very low. However, 6-MSA was still produced resulting in a final titer of 230 mg/L, which was substantially more than the about 100 mg/L produced during the glucose growth phase. Interestingly, when the concentration of ethanol was 63 g/L, which resulted from the fermentation with the initial glucose concentration of 200 g/L, the cells were totally inhibited, as a diauxic shift in biomass production and CO_2 production were not observed. This resulted in no change of the 6-MSA concentration, which stayed at 100 mg of 6-MSA/L throughout the rest of the fermentation.

The effects of the increasing ethanol concentration in the media on the specific growth rate, ethanol consumption rate and the 6-MSA yield are illustrated in Figure 5.3 where these parameters are plotted as function of the initial glucose concentration. Figure 5.3 reveals that even though the production of 6-MSA at the final stage increased from 250 up to 570 mg/L when the initial glucose concentration was 20 and 50 g/L, the 6-MSA yield on ethanol was actually decreased. Furthermore, the higher concentration of ethanol reduced the growth of the cells and the ethanol consumption rates, resulting in lower 6-MSA yield.

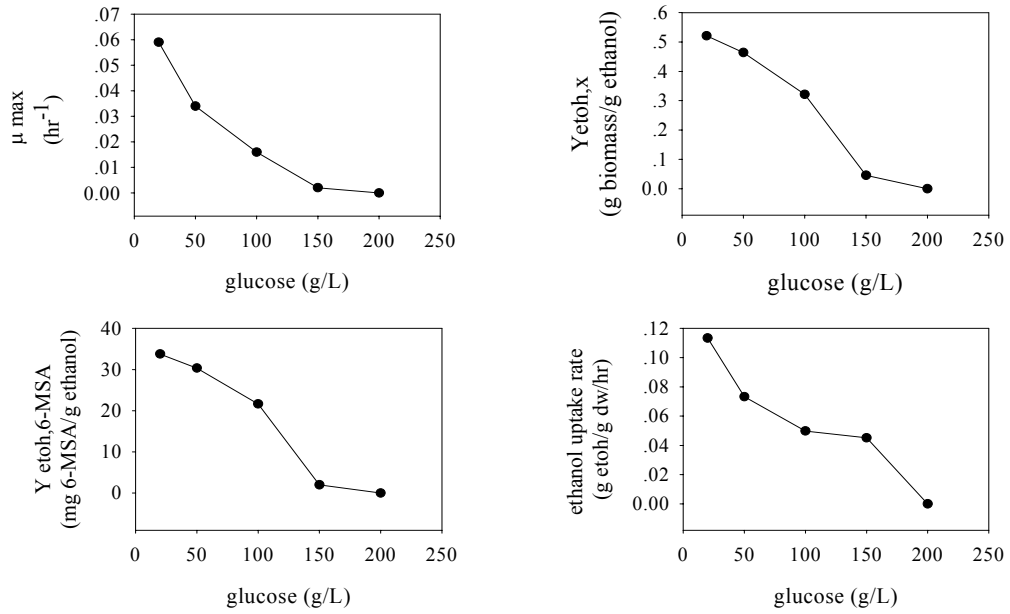


Figure 5.3. The changes of growth rate, ethanol uptake rate and product yields during the ethanol phase when cultivated *S. cerevisiae* IBT100083 in minimal media with different initial glucose concentrations.

5.4 Discussion

The idea of this study was originated from the previous observation that the yield of 6-MSA was substantially increased when *S. cerevisiae* carrying the 6-MSAS gene grew during the ethanol growth phase in aerobic cultivations. This made us believe that we could cultivate the cells in a medium with higher concentration of glucose (and hence also ethanol) in order to obtain higher amount of 6-MSA. We therefore cultivated the yeast cells in a glucose minimal medium at aerobic growth conditions with high initial concentrations of glucose as this will result in high concentrations of ethanol due to overflow metabolism towards ethanol in the cells. Furthermore, the cells will slowly adapt to the increasing ethanol concentration, and hence continue to consume the formed ethanol, which could be difficult if they were directly inoculated in media containing high ethanol concentrations.

During the batch cultivations in the first exponential phase where glucose was consumed, we found that the maximum specific growth rate of the cells, the yields of primary metabolites and 6-MSA on glucose decreased with high initial concentrations of glucose, especially when the initial glucose concentration was more than 100 g/L. The effect of high concentration of glucose on the growth and biomass formation was also studied by Ryu et al. (2000), who observed that when the glucose concentration in fed-batch cultivations increased above 400 g/L, the formation of biomass and the concentration of the product, erythritol, in the media were decreased. The inhibition on the growth of yeast cells by the high concentration of glucose made Overkamp et al. (2002) to search for a glucose tolerant strain in order to use this for high-level production of glycerol.

Moreover, the decrease in the specific growth rate at high glucose concentrations observed here was also according to the observations from other studies. Most of them revealed that the increase in the sugar concentration from 120 to 180 g/L decreased the growth and viability of the yeast cells (Casey and Ingledew, 1986; Xu et al., 1996; Bafrnacová et al., 1999; Ivorra et al., 1999), likely because of osmotic stress and a complex response that included rapid reduction in internal cell volume by water efflux (Meikle et al., 1988; Marechal and Gervais, 1994; Ivorra et al., 1999). Thomas et al. (1996) studied anabolic and catabolic pathway regulation at high glucose concentrations to understand its effect on cell metabolism and they established evidence that the synthesis of glycolytic enzymes as well as enzymes of the hexose monophosphate pathway, are regulated by glucose concentration. Furthermore, there are other possible explanations for the growth inhibition at high glucose concentrations, including the toxicity of some yeast sub-products, such as hexanoic, octanoic, and decanoic acids (Muñoz and Ingledew, 1990). However, Salmon (1989), Mauricio and Salmon (1992), and Salmon et al., (1993) proved that the principal factor limiting yeast metabolism at high glucose concentrations is an inhibition of sugar transport. Sugar transport in yeast has been widely studied. It is characterized by the presence of several transporters that have a distinct affinity to glucose (Ciriacy and Reifenberger, 1997). Those presenting high affinity to the substrate are subject to catabolic repression and are not detected during fermentations at high glucose

concentrations, indicating that these transporters are repressed or inhibited (Bisson et al., 1987).

Nevertheless, in the yeast fermentations, after glucose was exhausted, the ethanol that was formed during the use of glucose was used as another carbon source. As expected, the higher initial glucose concentration, the higher ethanol concentration was obtained. Even though ethanol has been found to be beneficial for the production of the polyketide 6-MSA in yeast, there seems to be a limited level of ethanol concentration that yeast can tolerate. This, however, depends on the yeast strains. In this study where the yeast IBT100083 was applied, we found that the specific growth rate, the yield of 6-MSA on ethanol and the specific ethanol uptake rate in the cells decreased when the concentration of ethanol in the medium increased. This was likely due to the fact that the ethanol at high concentrations inhibit essential cellular functions, giving rise to toxic effects especially on plasma membrane and cytosolic enzymes (Lopes and Sola-Penna, 2001). Furthermore, non-volatile compounds produced during the growth on ethanol can also hinder the growth of the cells. All of these effects could be the reasons why the high initial concentration of glucose, e.g. 100, 150 and 200 g/L, the growth of the yeast cells decreased and the yields of 6-MSA were less than when there was used lower initial glucose concentratin, e.g. 20 and 50 g/L.

In summary, *S. cerevisiae* IBT100083 was cultivated in glucose minimal media with different initial glucose concentrations ranging from 20 to 200 g/L and this resulted in different amount of ethanol formed. At low initial glucose concentrations (20 and 50 g/L), the yeast exhibited a high specific growth rate with high production and a high 6-MSA yield on glucose. The growth of yeast cells started to be inhibited when the initial glucose concentrations was higher than 50 g/L, resulting in lower production of 6-MSA and other primary metabolites, as well as the yields on both glucose and ethanol dropped. This was due to high osmotic stress caused by glucose and the toxic effects from ethanol and other organic compounds generated during the cultivations. Thus, when the initial glucose concentration was 200 g/L resulting in 63 g/L of ethanol there was slow growth on glucose and no growth on ethanol.

5.5 References

Bafrnaconá P, Smogrovicova D, Slavikova I, Patkova J, Domeny Z. 1999. Improvement of very high gravity ethanol fermentation by media supplementation using *Saccharomyces cerevisiae*. Biotechnol. Lett. 21: 337-341.

Bisson LF, Neigeborn L, Carlson M, Fraenkel DG. 1987. The SNF3 gene is required for high-affinity glucose transport in *Saccharomyces cerevisiae*. J. Bacteriol. 169: 1656-1662.

Casey GP, Ingledew WM. 1986. Ethanol tolerance in yeast. Crit. Rev. Microbiol. 13: 219-280.

Ciriacy M, Reifenberger M. 1997. Hexose transport. *In* Zimmermann FK, Entian K (eds.). Yeast sugar metabolism. Technomic Publishing Company, Lancaster, Pennsylvania.

Fiechter A, Fuhrmann GF, Käppeli O. 1981. Regulation of glucose metabolism in growing yeast cells Adv. Microb. Physiol. 22: 123-183.

Ivorra C, Pérez-Ortin J, Del Olmo M. 1999. An inverse correlation between stress resistance and stuck fermentation in wine yeast. A molecular study. Biotechnol. Bioeng. 64(6): 698-708.

Käppeli O. 1986. Regulation of carbon metabolism in *Saccharomyces cerevisiae* and related yeasts. Adv. Microb. Physiol. 28: 181-209.

Lopes DH, Sola-Penna M. 2001. Urea increases tolerance of yeast inorganic pyrophosphatase activity to ethanol: the other side of urea interaction with proteins. Arch. Biochem. Biophys. 394(1): 61-66.

Marechal PA, Gervais P. 1994. Yeast viability related to water potential variation: influence of the transient phase. Appl. Microbiol. Biotechnol. 42: 617-622.

Mauricio JC, Salmon JM. 1992. Apparent loss of sugar transport activity in *Saccharomyces cerevisiae* may mainly account for maximum ethanol production during alcoholic fermentation. Biotechnol. Lett. 14(7): 577-582.

Meikle AJ, Reed RH, Gadd GM. 1988. Osmotic adjustment of organic solutes in whole cells and protoplasts of *Saccharomyces cerevisiae*. J. Gen. Microbiol. 134: 3049-3060.

Muñoz E, Ingledew WM. 1990. Yeast hulls in wine fermentations. A review. J. Wine Res. 1(3): 197-209.

Overkamp KM, Bakker BM, Kötter P, Luttik MAH, van Dijken JP, Pronk JT. 2002. Metabolic engineering of glycerol production in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 68(6): 2814-2821.

Pronk JT, Steensma HY, van Dijken JP. 1996. Pyruvate metabolism in *Saccharomyces cerevisiae.* Appl. Environ. Microbiol. 55: 468-477.

Ryu Y-W, Park CY, Park JB, Kim S-Y, Seo J-H. 2000. Optimization of erythritol production by *Candida magnoliae* in fed-batch culture. J. Ind. Microbiol. Biotechnol. 25: 100-103.

Salmon JM. 1989. Effect of sugar transport inactivation in *Saccharomyces cerevisiae* on sluggish and stuck enological fermentations. Appl. Environ. Microbiol. 55(4): 953-958.

Salmon JM, Vincent O, Mauricio JC, Bely M, Barre P. 1993. Sugar transport inhibition and apparent loss of activity in *Saccharomyces cerevisiae* as a major limiting factor of enological fermentations. Am. J. Enol. Vitic. 44(1): 56-54.

Thomas K, Hynes SH, Ingledew WM. 1996. Effect of nitrogen limitation on synthesis of enzymes in *Saccharomyces cerevisiae* during fermentation of high concentration of carbohydrates. Biotechnol. Lett. 18(10): 1165-1168.

Xu P, Thomas A, Gilson CD. 1996. Combined use of three methods for high concentration ethanol production by *Saccharomyces cerevisiae*. Biotechnol. Lett. 18(2): 1439-1440.

Chapter 6

Transcriptional and Metabolic Responses to Heterologous Expression of a Polyketide Synthase in *S. cerevisiae*

Abstract

Natural producers of many interesting polyketides are difficult to cultivate and characterize, and it is therefore desirable to use more well characterized host cells for production of these natural products. *S. cerevisiae* is a suitable host for heterologous polyketide production of fungal polyketides as the knowledge of genomic and physiological properties of this organism is well established and *S. cerevisiae* is easy to handle in a fermentation process. Moreover, it has been reported that a high polyketide yields can be obtained with this organism. In order to make *S. cerevisiae* a general platform for production of polyketides, it is valuable to understand how the metabolism is regulated and affected by heterologous expression of a polyketide synthase. We therefore performed genome-wide transcription analysis and metabolic network analysis of *S. cerevisiae* carrying plasmids with the 6-methylsalicylic acid synthase (6-MSAS) gene from *Penicillium patulum* which is a model polyketide synthase, and a phosphopant-etheinyl transferase (PPTase) gene, *npgA* from *Aspergillus nidulans*. The recombinant and the reference strain, carrying empty plasmids, were grown in batch cultivations and cells for analysis were harvested from both the glucose and the ethanol growth phase. Furthermore, a strain with improved 6-MSA production, having over-expression of the gene *ACC1* encoding acetyl-CoA carboxylase was included in the study.

Evaluation of the transcriptome data using reporter metabolite and subnetwork structure analysis, to identify coordinated regulation of genes, revealed that the expression of 6-MSAS had a large effect on several parts of the metabolism. In particular amino acid metabolism was found to be significantly affected when comparing the strain carrying the 6-MSAS gene to the reference strain. However, also glycolysis and other parts of the carbohydrate metabolism, e.g. pathways leading to storage carbohydrates were significantly influenced as a result of the heterologous polyketide synthase expression. The most noteworthy change of expression for the improved strain with over-expression of the *ACC1* gene was observed for genes connected to fatty acid metabolism.

Furthermore, by using [1-^{13}C]-labeled glucose as carbon source and measuring metabolite labeling patterns, effects on the fluxes in the metabolic network due to 6-MSAS expression were investigated. It was shown that the flux towards acetyl-CoA was increased in the strains carrying the 6-MSAS gene, which corresponds to the amount of 6-MSA produced by the cell.

6.1 Introduction

Polyketides are secondary metabolites from bacteria, fungi, plants and animals which are derived from the polymerization of short fatty acids (most common are acetate, malonate or propionate) in a process similar to fatty acid synthesis. Polyketides are structurally a very diverse family of natural products with an extremely broad range of biological activities and pharmacological properties often use as antibiotics, antifungals, cytostatics, anticholesterolemics, antiparasitics, animal growth enhancers and natural insecticides. They represent the most medical important group of microbial metabolites with 20% of the top 10 selling drugs are polyketides or derived from polyketides.

6-Methylsalicylic acid (6-MSA) has been used as a model polyketide for studies on heterologous expression. It has earlier been demonstrated that *S. cerevisiae* carrying the 6-MSAS gene from *Penicillium patulum*, with the coexpression of a gene encoding phophopantetheinyl transferase (PPTase) is capable of producing significant amounts of 6-

MSA (Kealey et al., 1998; Wattanachaisaereekul et al., 2007). Furthermore, improved 6-MSA production has been demonstrated as a result of overexpression of *ACC1* encoding acetyl-CoA carboxylase, which is the enzyme catalyzing the conversion of acetyl-CoA into malonyl-CoA (Chapter 3). This was done by replacing the native promoter of *ACC1* with the strong, constitutive *TEF1* promoter. Physiological characterization of the improved strain has shown that *S. cerevisiae* is capable of producing 6-MSA with good yields, particularly when ethanol is the carbon source, resulting in a considerable drain of acetyl-CoA into 6-MSA. Thus, in this strain the flux of acetyl-CoA into 6-MSA production was as high as about 40% of the total drain of acetyl-CoA, with the remaining acetyl-CoA consumption mainly going into lipid biosynthesis (Chapter 3).

The purpose of the present study was to gain a more detailed understanding of the effects on metabolism when yeast is producing 6-MSA, and also evaluate the effect of draining a large part of the acetyl-CoA pool into 6-MSA production. Hence, genome-wide transcription analysis and metabolic network analysis were applied on three different strains during aerobic growth in batch cultivations and samples were harvested both from the glucose and the ethanol phase. The *S. cerevisiae* strains investigated were a reference strain, carrying empty plasmids (IBT100081) and two different strains expressing the gene encoding 6-MSA synthase (6-MSAS) from *P. patulum* with co-expression of a PPTase from *A. nidulans* (IBT100082, IBT100083) (Chapter 3). One of the 6-MSAS expressing strains also has the *ACC1* under the control of the strong, constitutive *TEF1* promoter (IBT100083).

6.2 Materials and methods

6.2.1 Strains

The three strains of *S. cerevisiae* used in this study have been described in Chapter 3 and they are listed in Table 6.1.

Table 6.1. List of strains used in this work.

Strains	Characteristics
IBT100081 (reference strain)	CEN.PK 113-9D carrying empty plasmids pRS424 and pRS426
IBT100082	CEN.PK 113-9D carrying plasmid with 6-MSAS gene from *P. patulum* (pRS426CTMSA-PP) and plasmid with the *npgA* encoding enzyme PPTase from *A. nidulans* (pRS424CTnpgA)
IBT100083	CEN.PK 113-9D which the promoter of *ACC1* is changed to *TEF1* promoter and carrying plasmid with 6-MSAS gene from *P. patulum* (pRS426CTMSA-PP) and plasmid with the *npgA* encoding enzyme PPTase from *A. nidulans* (pRS424CTnpgA)

6.2.2 Bioreactors and cultivation conditions

The strains were cultivated in 100 mL medium in 500 mL baffled shake flasks for propagation. The medium contained: glucose 10 g/L, $(NH_4)_2SO_4$ 7.5 g/L, KH_2PO_4 14.4 g/L, $MgSO_4.7H_2O$ 0.5 g/L, 0.05 mL/L Antifoam 298 (Sigma-Aldrich, St. Louse, MO, USA) and 2 mL/L trace metal solution ($FeSO_4.7H_2O$ 3 g/L, $ZnSO_4.7H_2O$ 4.5 g/L, $CaCl_2.6H_2O$ 4.5 g/L, $MnCl_2.2H_2O$ 0.84 g/L, $CoCl_2.6H_2O$ 0.3 g/L, $CuSO_4.5H_2O$ 0.3 g/L, $Na_2MoO_4.2H_2O$ 0.4 g/L, H_3BO_3 1 g/L, KI 0.1 g/L, and $Na_2EDTA.2H_2O$ 15 g/L), and 2 mL/L vitamin solution (d-biotin 50 mg/L, Ca-pantothenate 1 g/L, thiamin-HCl 1 g/L, pyridoxin-HCl 1 g/L, nicotinic acid 1 g/L, p-aminobenzoic acid 0.2 g/L, and m-inositol 12.5 g/L). pH of the medium was adjusted to 6.5 by 2M NaOH prior to autoclaving. The cultures were incubated at 30 ^oC, with shaking at 150 rpm (model 3033, GFL, Burgwedel, Germany). When the absorbance (A_{600}) reached 13 (2.3 g biomass/L), 0.1 mL of the culture was used as inoculum for the batch cultivations.

Batch cultivations for transcription analysis with two duplicates were carried out in bioreactors (Biostat Braun Biotech International GmbH, Mulsungen, Germany) with a working volume of 2 L. The medium consisted of glucose 20 g/L, $(NH_4)_2SO_4$ 10 g/L, KH_2PO_4 3 g/L, $MgSO_4.7H_2O$ 1 g/L with 0.05 mL/L Antifoam 298, 2 mL/L of trace metal

solution and 2 mL/L of vitamin solution. The temperature of the cultivations was kept at 30^{o}C, and the pH was automatically controlled at 5.0 by addition of 2M KOH. The agitation was 500 rpm and the bioreactors were aerated with 2 L air/min.

Batch cultivations for flux analysis were carried out in the same kind of bioreactor with a working volume of 0.6 L. The medium composition was the same as used in the transcription analysis experiment except that half of the glucose used was labeled ([1-^{13}C]-labeled glucose; ISOTEC^{TM}, Miamisburg, OH, USA). The cultivations were controlled at 30^{o}C, pH 5.0, with an agitation of 500 rpm and aerated with 0.6 L air/min.

6.2.3 Sampling

For gene expression analysis, 20 mL samples were withdrawn in mid-exponential phase, one sample in the glucose phase and another one in the ethanol phase. The sample was rapidly poured into a 50 mL Falcon tube containing 30 mL crushed ice, mixed and pelleted at 0^{o}C, 4000 rpm for 5 minutes. The supernatant was discarded and the tube containing the cells was immediately put in liquid nitrogen, and stored at -80°C until further analysis.

For flux analysis, the samples were taken throughout the cultivation period, 6 samples during the first exponential glucose phase, and another 4 samples in the ethanol phase. The sample volume withdrawn was adjusted at each time point to give around 50 mg wet weight of biomass that is required for the analysis. The samples were immediately centrifuged at 0^{o}C, 4000 rpm for 5 minutes. After the centrifugation, the supernatant was removed and kept at -20^{o}C for analysis of glucose, primary metabolites and 6-MSA. The cell pellet was frozen instantaneously in liquid nitrogen and stored at -80°C before performing the biomass hydrolysis. Each time a sample for flux analysis was withdrawn 5 mL of the culture was also taken for analysis of biomass dry weight.

6.2.4 Biomass measurement

Biomass dry weight was determined by filtering 5 mL fermentation broth, through a dried, pre-weighed nitrocellulose filter (Gelman Science, Ann Arbor, MI) with a pore size 0.45 μm. The residue was washed twice with distilled water. The filter was dried to constant weight in a microwave oven at 150 W for 10 minutes, cooled down in a desiccator, and the weight gain was measured.

6.2.5 Analysis of glucose, primary metabolites and 6-MSA

Fermentation samples were centrifuged and the supernatant was taken for later analysis of substrate and products. Glucose and primary metabolites was analyzed by high-performance liquid chromatography (Dionex-HPLC; Sunnyvale, CA), equipped with an Aminex HPX-87H ion exclusion column, 300 mm × 7.8 mm (Bio-Rad Laboratories, Hercules, CA), which was operated at 60^{o}C, and a flow rate of 0.6 mL/min of 5 mM H_2SO_4 using a refractive index detector (Shodex RI-71, Tokyo, Japan) and a UV detector (UVD340S, Dionex). 6-MSA was quantified from the filtrate by HPLC (Agilent 1100 series) equipped with Luna C18(2) column, 150 × 4.60 mm with 5 micron of the porous silica particles to which the C18 phase is bonded, using a gradient of 50 ppm trifluoroacetic acid (TFA) in milliQ water (solvent A) and 50 ppm TFA in acetonitrile (solvent B) at a flow rate of 0.5 mL/min. The gradient of the solvents was 20% to 60% B in 10 minutes, then 20% B for the next 2 minutes.

6.2.6 Extraction of total RNA

Total RNA extraction was performed using a FastRNA®, Red Kit (Qbiogene, Illkirch Cedex, France). After thawing the samples on ice, the phenol reagent was added directly on the cells and added up to 1.2 mL into the tube containing glass bead. The mixture was shaken for 40 seconds at level 6 (Fast Prep, FP120, Savant Instrument Inc., Holbrook, NY, USA). The cell debris was pelleted and discarded. The supernatant was extracted three times by chloroform, 350, 300 and 200 μL, respectively. The top phase (phenol reagent phase) was transferred to a new tube and 500 μL of 80% cold ethanol at -20^{o}C was added.

The samples were kept at -20 ^{o}C for 2 hours to pellet the RNA. The RNA pellet was washed by 80% cold ethanol and air dried. The dry pellet was dissolved in RNAse free water (ultraPURE, GIBCO^{TM}, Grand Island, NY, USA). The quantity and quality of the extracted RNA was determined by DNAquant II (Pharmacia Biotech, Cambridge, England) and 2100 bioanalyzer using the RNA 6000 Nano Assay (Agilent Technologies, Inc., Santa Clara, CA, USA). The RNA was stored at -80^{o}C until further processing.

6.2.7 Probe preparation and hybridization

cDNA synthesis, cRNA synthesis and biotin labeling of antisense cRNA were carried out according to Affymetrix GeneChip® manual. 15 μg of fragmented, labeled cRNA was hybridized to Yeast Genome 2.0 oligonucleotide probe arrays (Affymetrix, Santa Clara, CA, USA) as described in the Affymetrix GeneChip® manual for array processing. Washing and staining of arrays were performed using the GeneChip Fluidics Station 450 and scanning with the Agilent Gene Array Scanner (Affymetrix).

6.2.8 Whole genome transcription data analysis

The acquisition and quantification of array images as primary data analysis were performed using the GeneChip® Operation Software (Affymetrix) v.1.1, resulting in the raw data CEL files. The individual chips were scaled to the global level of hybridization and normalized to the same median intensity using the qspline method, thereby making them comparable to each other (Workman et al., 2002). For each gene an expression index was calculated based on the probes by using the Li-Wong Model-Based Expression index (PM-only model) (Li and Wong, 2001). All array data were subsequently analyzed for the principal component analysis to discover any grouping in the chips.

Statistical analysis was applied to determine the difference of the gene expression between the categories of replicated experiments. A Student t-test was employed for different pair-wise comparisons in the glucose phase (IBT100082 versus IBT100081) and in the ethanol phase (IBT100082 versus IBT100081, and IBT100083 versus IBT100082). Furthermore the comparison of the strain IBT100082 in glucose phase versus ethanol phase was also

performed. Gene description and annotations were found in the Saccharomyces Genome Database (SGD; http://www.yeastgenome.org) by using SGD Gene Ontology Term Finder.

The reporter metabolites (metabolites around which the most significant changes in the transcription occur) and highly correlated subnetwork (sets of connected genes with significant and coordinated transcriptional response to a perturbation) as described by Patil and Nielsen (2005), were also identified for each pair-wise comparison in order to be able to extract more information about the effects on metabolism and regulation due to the heterologous expression of 6-MSA compared to the reference strain in both the glucose and the ethanol phase.

6.2.9 GC-MS analysis and fractional labeling of intracellular metabolites

In order to quantify the intracellular fluxes, a combination of metabolite balancing and isotope labeling experiments were applied. The labeling experiments were performed by feeding the cultures with labeled glucose, allowing the labeled carbon atoms to distribute over the metabolic network, and subsequently analyse the [1-^{13}C]-labeled distributions of intracellular compounds. The [1-^{13}C]-labeling of the intracellular metabolites were determined by an indirect method based on measurements of the labeling pattern of the proteinogenic amino acids using GC-MS (Christensen and Nielsen, 1999). For GC-MS analysis, derivatization of amino acids is necessary. The amino acid derivatization applied in this study, originally developed by Christensen and Nielsen (1999) and later refined by Bapat et al. (unpublished data), starts with the hydrolysis of biomass for preparing the derivatives of amino acids, solid phase extraction (SPE) of amino acids from biomass hydrolyzate using Dowex cation exchange resin, subsequently by performing three methods 1) Ethylcholoformate derivatization (ECF) for Ala116, Asp188, Gly175, Ile158, Leu158, Lys156, Phe192, Pro142, Ser132, Ser175, Val144; 2) N,N-Dimethylformamide dimethyl acetal derivatization (DMFDMA) for Ala99, Ala158, Asp115, Asp216, Glu143, Glu230, Gly85, Gly144, Phe143, Val127, Val143, Val186; 3) Pyridine-Acetic acid anhydride derivatization (PAA) for Glc331. After the derivatization, each amino acid

isotopomer was detected by GC-MS according to the protocol described by Christensen and Nielsen (1999) which was subsequently developed by Bapat et al. (unpublished data).

The raw data from GC-MS analysis were normalized using the summed fractional labeling (SFL) concept described elsewhere by Christensen and Nielsen (2000). This calculated SFL includes the natural labeling present in the carbon derived from the derivatization. In order to obtain the true SFL, the SFL value was corrected for the natural labeling of the carbon atoms derived from the derivatization agent and for the natural labeling of oxygen, hydrogen and nitrogen content of the amino acids (Wittmann and Heinzle, 1999). The consistency of the labeling measurements was evaluated by comparing the SFLs of a specific fragment, which were obtained by different methods, or by comparing different fragments of the same molecule resulting from a particular derivatization method.

6.2.10 Flux quantifications

The *in vivo* metabolic fluxes were estimated by combining metabolite balancing and isotope balancing. Metabolite balancing relies on the stoichiometry of the reactions that comprise the metabolic network and on a set of measured fluxes, giving rise to a linear system of equations with respect to the metabolic fluxes, and the isotope balancing is carried out on a carbon atom basis or the fraction of labeled carbon atoms in a given position of a metabolite. The combination of metabolite balancing and isotope balancing enabled the metabolic flux calculation without the need for the assumptions of co-factor requirement. The flux calculation in this study was accomplished by using C13 software which was previously developed by Christiansen and Nielsen (2000), and further extended by Grotkjær and David (2005). The mathematical framework adopted in this software was the one developed by Wiechert and de Graaf (1997).

6.3 Results and discussion

6.3.1 Physiological characterization

The strains *S. cerevisiae* IBT100081, 100082 and 100083 were cultivated in duplicates in glucose minimal media (20 g/L of glucose) in batch cultivations and the observed product yields from this study (Table 6.2) were in agreement with those previously described in Chapter 3. A summary of the batch cultivation data is given in the following to provide a background for interpretation of the transcription data and flux analysis. In the first exponential phase where glucose was used as a carbon source, the maximum specific growth rate (μ_{max}) of the reference strain (IBT100081) was 0.33 hr^{-1}, whereas those of IBT100082 and IBT100083 were 0.26 and 0.27 hr^{-1}, respectively. The 6-MSA yield of IBT100082 was 1.4 mg/g glucose whereas the yield of IBT100083, having the native *ACC1* promoter exchanged with the strong constitutive promoter *TEF1*, was 2.0 mg/g glucose. These results indicate an improved supply of the 6-MSA precursor malonyl-CoA due to the stronger expression of *ACC1* and hence a higher production with IBT100083. No 6-MSA was produced by the reference strain IB100081 as expected as it carried the empty plasmids without the heterologous PKS gene.

In the second exponential phase where the ethanol generated during growth on glucose was used as carbon source, the μ_{max} of IBT100081, 100082 and 100083 were 0.06, 0.05 and 0.06 hr^{-1}, respectively (Table 6.2). Ethanol has earlier been observed to be a better carbon source for polyketide production than glucose (Wattanachaisaereekul et al., 2007) and in line with this the 6-MSA yield was higher for both IBT100082 and 100083 in the ethanol phase, 19 and 34 mg/g ethanol, respectively.

Table 6.2. The maximum specific growth rate and product yields of the three different strains in batch cultivations. Data were calculated both for the glucose and the ethanol growth phase.

Growth rate and product yields	Strains		
	IBT100081	IBT100082	IBT100083
Glucose phase			
μ_{max} (hr^{-1})	0.33 ± 0.01	0.26 ± 0.01	0.27 ± 0.00
Y_{se} (mg ethanol/g glucose)	359.2 ± 3.7	358.4 ± 7.8	370.8 ± 3.7
Y_{sg} (mg glycerol/g glucose)	47.5 ± 0.3	46.2 ± 4.9	44.6 ± 0.9
Y_{sa} (mg acetate/g glucose)	9.9 ± 0.2	10.7 ± 0.9	12.4 ± 0.7
Y_{sm} (mg 6-MSA/g glucose)	0.0 ± 0.0	1.4 ± 0.1	2.0 ± 0.0
Y_{sx} (mg biomass/g glucose)	125.9 ± 2.2	112.5 ± 1.0	114.3 ± 0.4
Ethanol phase			
μ_{max} (hr^{-1})	0.063 ± 0.001	0.052 ± 0.004	0.059 ± 0.001
Y_{sm} (mg 6-MSA/g ethanol)	0.0 ± 0.0	19.3 ± 2.7	33.8 ± 0.1
Y_{sx} (mg biomass/g ethanol)	572.4 ± 4.9	494.9 ± 9.2	520.7 ± 5.3

6.3.2 Transcription data and analysis

Cells were harvested for transcription analysis both in the mid-exponential growth phase on glucose (IBT100081 and IBT100082) and in the mid-exponential part of the ethanol phase (IBT100081, IBT100082 and IBT100083). At each time point two cell samples were harvested, and consequently 10 arrays were being processed in order to assess the variance in expression in each gene between biological duplicates. The expression data were normalized using dChip which is based on the algorithm described by Li-Wong. From the yeast 2.0 genome Affymetrix array, the expression values for 5,814 annotated ORFs were identified. From this, all genes with an absent call in all arrays were removed. This pre-processing part with a file of normalized expression value resulted in data for 5,636 genes.

A principal component analysis (PCA) was performed for all the arrays based on 5,636 genes to evaluate the consistency of data (Figure 6.1). There is a clear separation in strains

in two dimensions, and in particular the PC2 separates the strains whereas samples from the glucose and the ethanol phase were separated by the PC1. A good repeatability of the experiments was obtained as replicates were grouping close together when visualized by PCA.

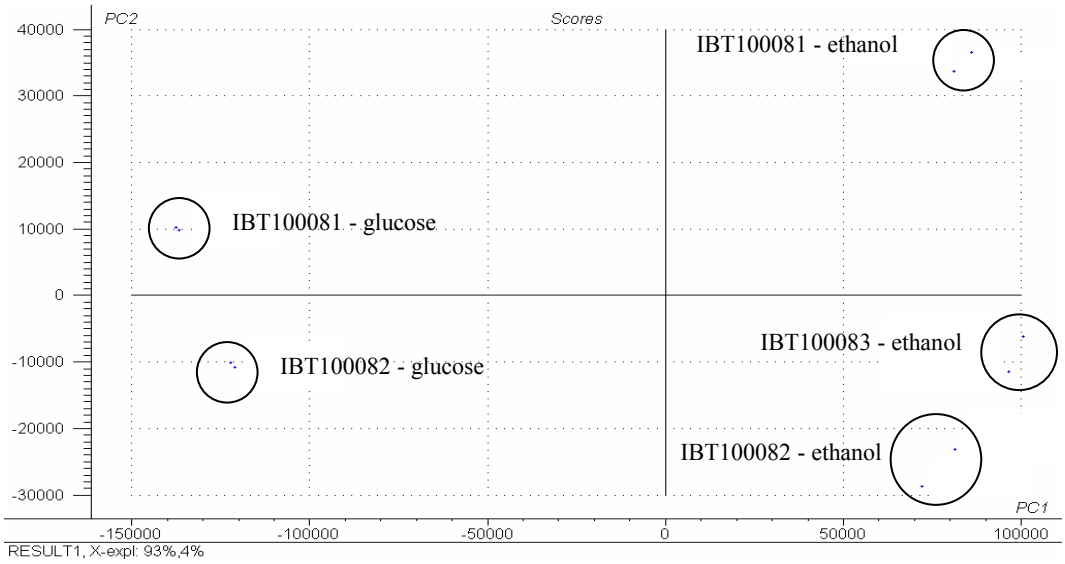


Figure 6.1. Principal component analysis of the three different strains, illustrating that the replicates of the experiments were grouping close together.

For the 5,636 ORFs, which were measured to be expressed under at least one of the time-points in at least one of the strains, pair-wise statistical comparisons were performed to identify significant changes in gene expression using the two-tails, null hypothesis with unequal variance t-test (Table 6.3). To assess the significant genes, a p-value cut-off of 0.05 was established for the comparison of IBT100082 vs IBT100081 in the glucose and the ethanol phase, respectively, and for IBT100083 vs IBT100082 in the ethanol phase. Using a smaller cut-off result in too few genes and further analysis is therefore hampered. However, in the comparison of IBT100082 in glucose versus ethanol phase, a p-value cut-off of 0.01 gave sufficient number of genes for further analysis.

Table 6.3. Total number of significant genes in the different pair-wise comparisons between the categories.

Cut-off (confidence level)	Number of significant genes (up-/down-regulated)			
	Glucose phase	Ethanol phase		IBT100082
	IBT100082 vs IBT100081	IBT100082 vs IBT100081	IBT100083 vs IBT100082	Ethanol phase vs Glucose phase
0.05 (95%)	777 (600/177)	717 (322/395)	795 (346/449)	2126 (1388/1224)
0.01 (99%)	173 (141/32)	136 (72/64)	129 (53/76)	935 (499/436)
0.001 (99.9%)	10 (9/1)	9 (4/5)	7 (3/4)	155 (88/67)

When comparing IBT100082 to IBT100081 in the glucose phase, 777 genes (600 up-regulated, 177 down-regulated) were identified as having significant changed mRNA levels. From these 777 genes, only 52 genes displayed a fold change of more than 2 fold. All 52 genes were up-regulated due to expression of the gene encoding 6-MSAS (strain IBT100082).

When the t-test was applied to the transcriptome data from cells harvested in the ethanol phase, a pair-wise comparison between IBT100082 and IBT100081 resulted in 717 genes that were significantly changed (322 up-regulated and 395 down-regulated) at a p-value cut-off of 0.05. Among the 717 genes, there were 54 genes (40 up-regulated and 14 down-regulated) with a more than 2 fold change. In the comparison of IBT100083 versus IBT100082, 795 significantly changed genes (346 up-regulated and 449 down-regulated) were obtained of which 26 genes (17 up-regulated and 9 down-regulated) were having an absolute fold change of more than 2. In the comparison of only IBT100082 in glucose and ethanol phase, which p-value cut-off of 0.01 was chosen, 935 significant genes (499 up-regulated and 436 down-regulated) was obtained which 103 genes (96 up-regulated and 7 down-regulated) were changed more than 2 fold. The distribution of the genes with significantly changed expression in the genome and the gene ontology (GO) for each comparison in both the glucose and the ethanol phase can be found in Appendix A.

6.3.3 Identification of reporter metabolites and metabolic subnetworks

For further understanding of the effect on gene transcription of expression of a polyketides synthase in *S. cerevisiae*, the algorithm developed by Patil and Nielsen (2005) was applied. This algorithm identifies subsets of genes with significant and coordinated transcriptional response connected via common metabolites, the so-called reporter metabolites. The identification of reporter metabolites is based on connecting the gene expression data with the metabolic network information from the genome-scale model of *S. cerevisiae* (Förster et al., 2003), which includes the genes encoding enzymes. Reporter metabolite analysis locates the parts of the metabolism that are most affected by a certain perturbation, hence enabling identification of commonly regulated genes.

The identification of highly correlated subnetworks was carried out using the p-values from the pair-wise comparisons as input data in combination with information from the whole reaction set in the genome-scale model of *S. cerevisiae*. Within the identified subnetworks, additional subnetwork structures were searched by the repetition of the algorithm over the subnetwork previously obtained, resulting in more robust and smaller subnetworks (Patil and Nielsen, 2005). The reporter metabolites and the enzymes comprising in the subnetwork for each of pair-wise comparisons between each strain both in glucose and ethanol phase are illustrated in Figure 6.2-6.5.

6.3.4 Global effect due to 6-MSA production

For production of 6-MSA, it has been proved that a higher productivity could be achieved in the ethanol phase than in the glucose phase, therefore gene expression for both phases in the 6-MSA production strain (IBT100082) was investigated. It is interesting to note that genes involved in the formation of acetyl-CoA, i.e. *ACS1* and *ACS2*, were up-regulated when ethanol was the carbon source. This is in agreement with the observations of others that a larger cytosolic acetyl-CoA pool is available during C2 carbon source utilization compared to glucose (Seker et al., 2005; van den Berg et al., 1996, 1998) and the improved 6-MSA production on ethanol may well be due to an increased acetyl-CoA availability. Also the reporter metabolite analysis points towards changes in the same area

of metabolism as acetate came out as a reporter metabolite in addition to the TCA metabolites oxaloacetate, malate and succinyl-CoA (Figure 6.2).

When comparing the expression patterns during growth on ethanol for the reference strain and the 6-MSA producing strain, IBT100081 and IBT100082, respectively, reporter metabolites related to the carbohydrate metabolism could be identified, with glycogen as the most significant one. There are four genes encoding neighboring enzymes around glycogen namely *GLC3*, *GPH1*, *GSY1* and *GSY2*. From the expression data, we found that *GLC3* was down-regulated, but *GPH1*, *GSY1* and *GSY2* all were up-regulated in the strain IBT100082 in the ethanol phase. The up-regulation of the two glycogen synthase genes (*GSY1* and *GSY2*) is in agreement with their description as being induced by glucose limitation, environmental stress by carrying a foreign gene, nitrogen starvation and entry into the stationary phase. All these reasons make glycogen as a top-scored reporter metabolite. We also found that the enzymes encoded by *URA3*, *URA10*, *TKL2*, *ABZ1*, *TRP1*, *TRP4*, *MET2*, *AAT1*, *ARG1*, *ARG8* and *GAD1* were included in the subnetwork (Figure 6.3), indicating that not only the glycogen synthases but also other enzymes relevant to carbohydrate and amino acid metabolism were influenced by 6-MSA biosynthesis.

6.3.5 Global effect on amino acid metabolism

From the statistical comparison of expression data using t-test, it was seen that several of the genes related to amino acid metabolism were significantly up-regulated in connection with 6-MSA biosynthesis. There were totally 38 significantly changed genes with 30 up-regulated genes and 8 down regulated genes related to amino acid metabolism, and the neighboring enzymes around L-alanine were also presented in these 38 genes. In agreement, the analysis of reporter metabolites also revealed that there were considerable effects on the amino acid metabolism when comparing the 6-MSA producing strain (IBT100082) to the reference strain (IBT100081) cultivated in glucose. L-alanine was found to be a top-scored reporter metabolite. This means that the expression of the genes around L-alanine was the most significantly changed. Moreover, several enzymes

involved in the amino acid metabolism were included in the subnetwork for this comparison (Figure 6.4), confirming that amino acid metabolism has been significantly affected by the presence of 6-MSAS when cultivating the strains on glucose.

6.3.6 Global effect on fatty acid metabolism

Changes in gene expression between the strain IBT100083 and IBT100082 in the ethanol phase were investigated in order to study the effect on the yeast metabolism after the promoter of *ACC1* has been replaced by the strong, constitutive *TEF1* promoter. Reporter metabolite analysis revealed that there were major effects on the lipid metabolism after the change of the *ACC1* promoter, episterol, a precursor in ergosterol biosynthesis, being the most significant reporter metabolite. Furthermore, other reporter metabolites were pyridoxine (vitamin B6) and derivatives thereof, which are connected to lipid metabolism as well (Figure 6.5).

The comparison of gene expression for the strain IBT100083 with IBT100082 in the ethanol phase by t-test also demonstrated drastic influence on lipid metabolism. There were totally 37 genes with significant changes in expression, 23 genes were up-regulated and the rest down-regulated. The neighboring enzymes of the identified reporter metabolites were found among these 37 genes, indicating the consistency of the result from expression data analysis with different methods. Additionally, the analysis of metabolic subnetworks for these two strains in the ethanol phase (Figure 6.5) displayed the connection of the enzymes in this subnetwork with episterol, and other metabolites in fatty acid biosynthesis as well as the pyridoxine biosynthesis, since the enzymes encoded by *ERG2, ERG3, PDX3, ACS1*, and *ACC1* were found to be co-regulated and significantly changed. This makes sense since the expression of *ACC1*, which encodes a key enzyme in fatty acid biosynthesis, had been changed. The observed changes in the lipid metabolism and especially in the expression of *ACC1* were expected. Interestingly, glutamate dehydrogenase (Gdh1) was found as a significant enzyme in the subnetwork structure when comparing IBT100083 to IBT100082 in the ethanol phase. Since the gene that encodes for Gdh1 was down-regulated in the strain IBT100083, this indicates that less

NADPH was consumed, and consequently, that there might be more NADPH available in the cytosol to be used as precursor in the biosynthesis of 6-MSA.

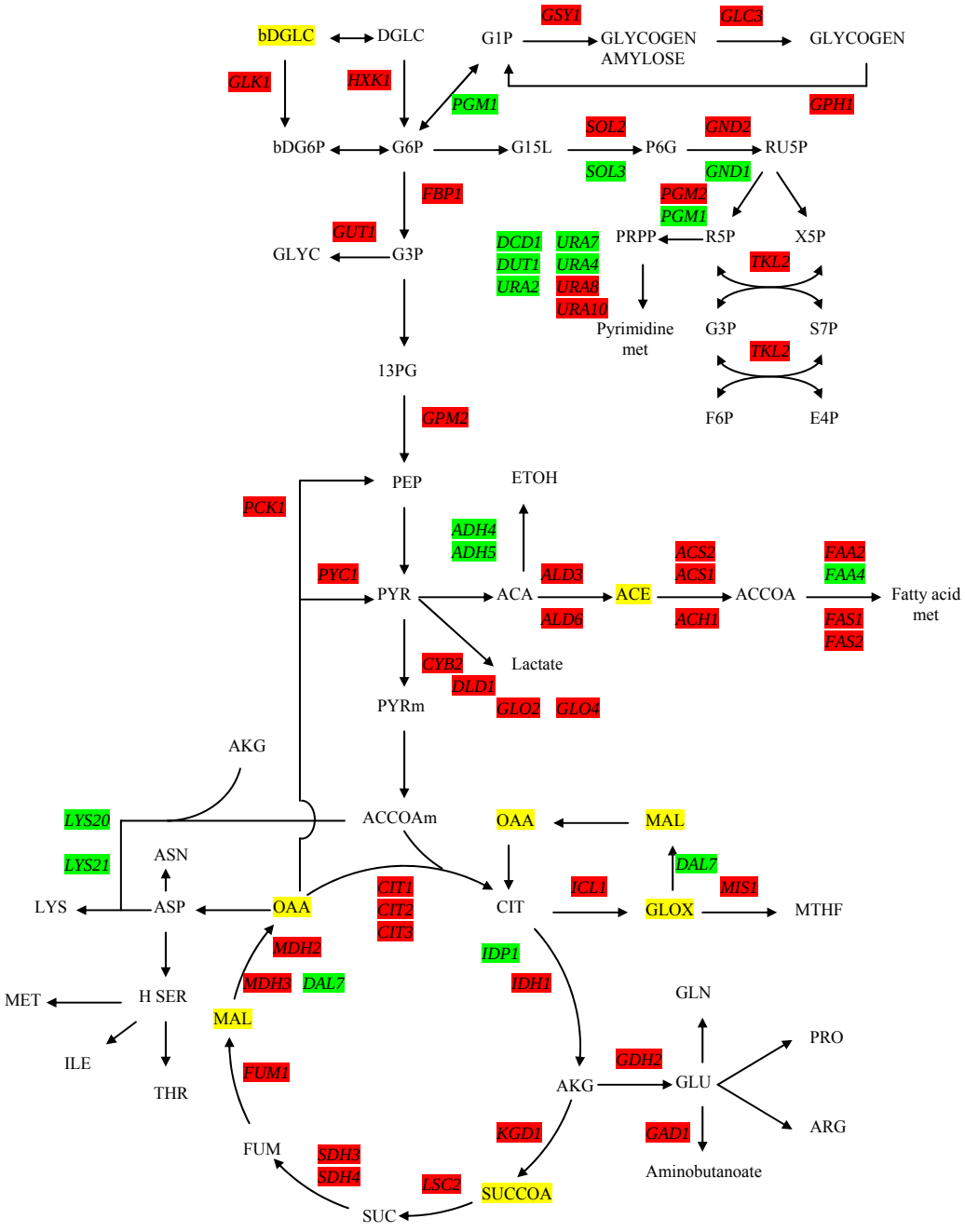


Figure 6.2. The reporter metabolites (highlight in yellow) and correlated subnetwork (red for up-regulated genes and green for down-regulated genes) when compared the strain IBT100082 in ethanol phase versus the same strain in glucose phase.

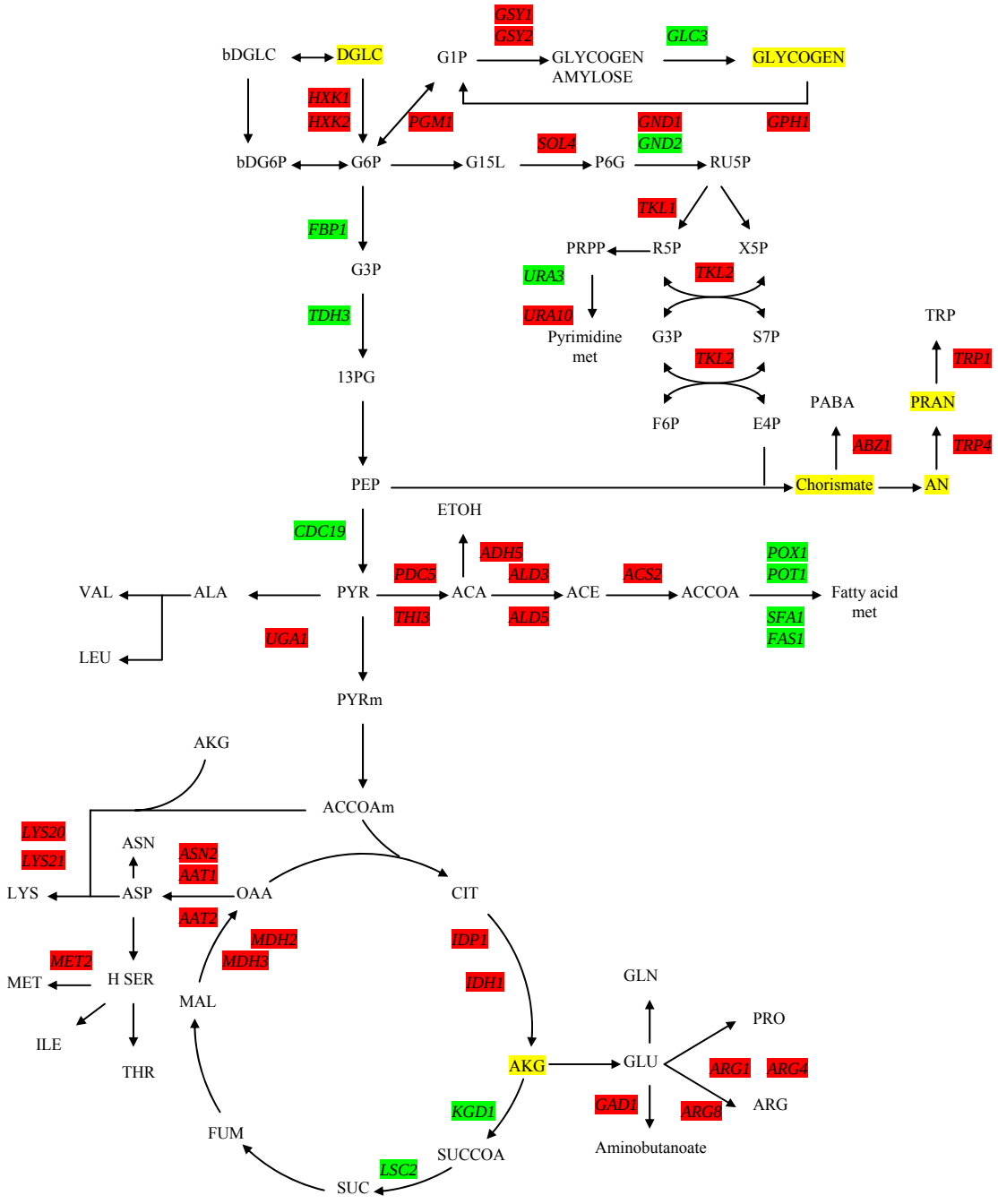


Figure 6.3. The reporter metabolites (highlight in yellow) and correlated subnetwork (red for up-regulated genes and green for down-regulated genes) when compared the strain IBT100082 versus IBT100081 in ethanol phase.

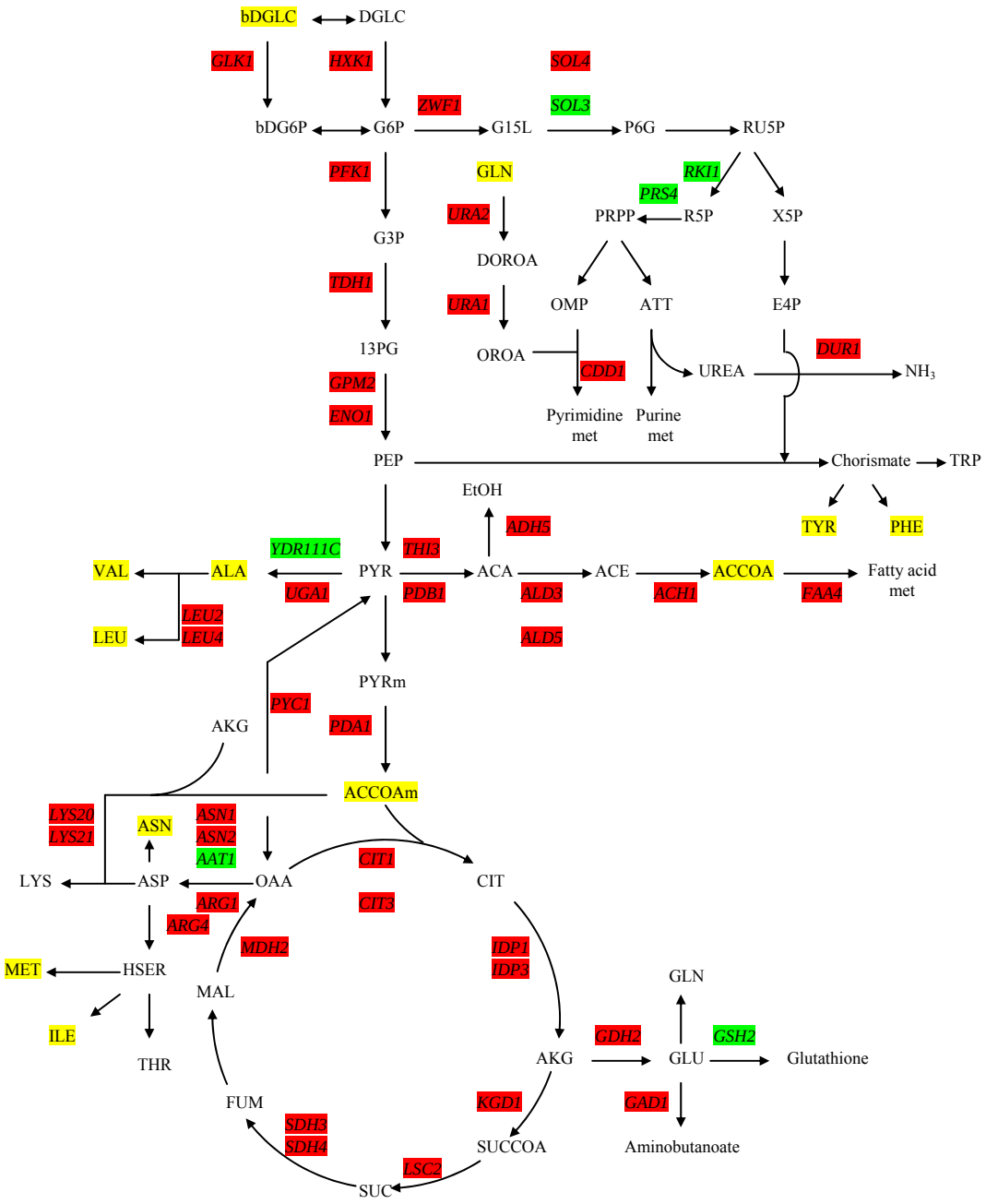


Figure 6.4. The reporter metabolites (highlight in yellow) and correlated subnetwork (red for up-regulated genes and green for down-regulated genes) when compared the strain IBT100082 versus IBT100081 in glucose phase, indicating the global effect on amino acid metabolism.

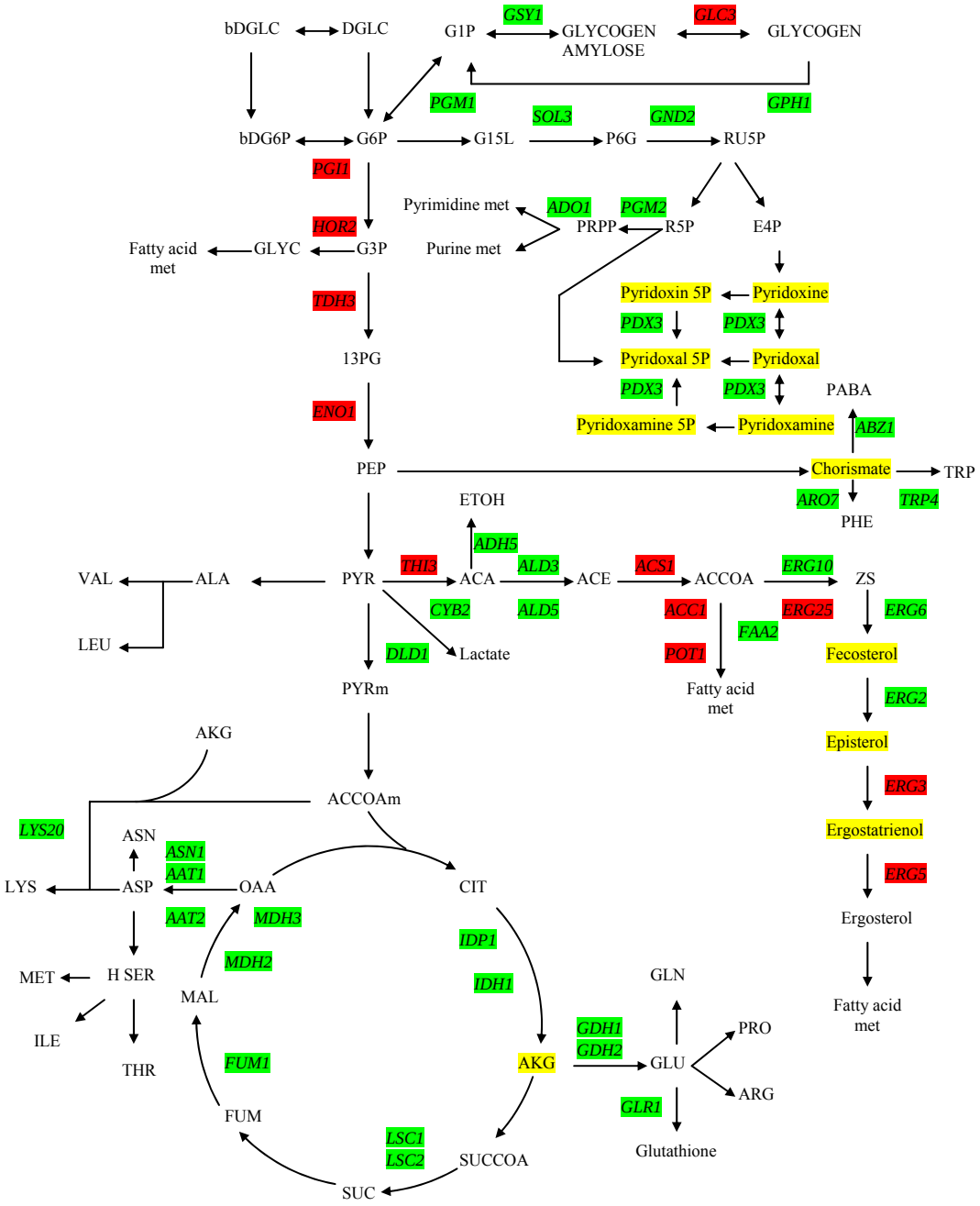


Figure 6.5. The reporter metabolites (highlight in yellow) and correlated subnetwork (red for up-regulated genes and green for down-regulated genes) when compared the strain IBT100083 versus IBT100082 in ethanol phase, indicating the global effect on fatty acid metabolism.

6.3.7 Summed fractional labeling

In order to characterize the metabolism and analyze the effect when there is 6-MSAS production, labeling experiment were performed as this allows for a quantitative as well as a qualitative study of the metabolism. Here [1-^{13}C]-labeled glucose was used as a carbon source for the growth of the strains IBT100081, 100082 and 100083 in batch cultivations.

Summed fractional labeling (SFL) of intracellular metabolite fragments of the reference strain and the strain carrying 6-MSAS from *P. patulum* were calculated. The experiments were carried out in the aerobic batch cultivations in which six samples were periodically taken from the fermenters during the first and second exponential phase where glucose and ethanol were consumed as carbon sources. The incorporation of label into some of the analyzed metabolite fragments from the *S. cerevisiae* carrying 6-MSAS from *P. patulum* and with over-expression of *ACC1* (IBT100083) is shown in Figure 6.6 with the concentration of CO_2 throughout the cultivation period, illustrating the first and second exponential and time to take samples (figure not shown for other strains).

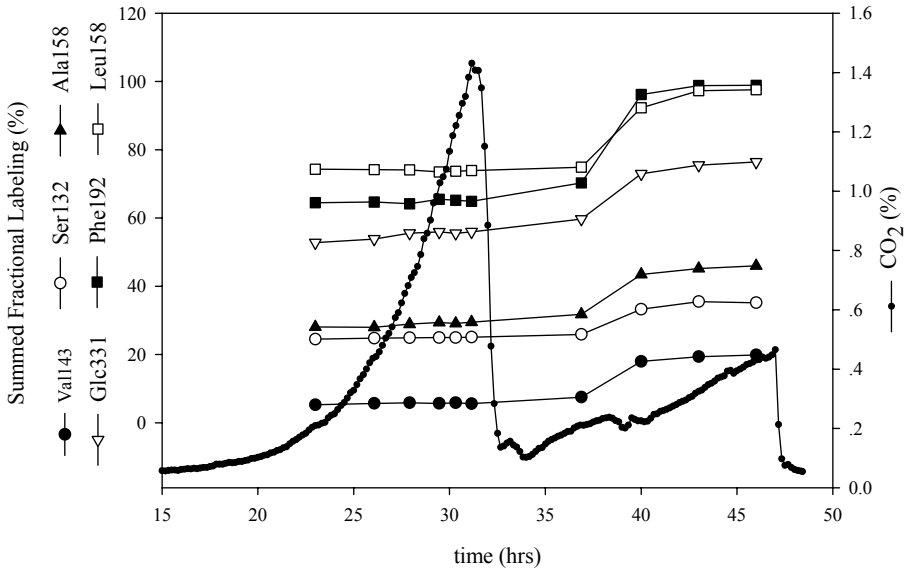


Figure 6.6. Summed fractional labeling of some selected intracellular metabolites of the strain IBT100083 during the batch cultivation. The CO_2 concentration was also shown to illustrate the glucose and ethanol phase.

From Figure 6.6 it can be seen that the SFLs calculated from the first six samples in the first exponential phase are constant which means that the isotopic pseudo steady state is achieved at this point. In contrast, SFL of the fragments are increased in the ethanol phase. The SFL of each fragments in the glucose phase for the three different strains were averaged and the data are listed in Table 6.4.

Table 6.4. Measured summed fractional labelings (SFLs) of the different strains from the first six samples in the first exponential phase where the glucose was consumed in the batch cultivations

Meta-bolite	m/z	Reported precursor		Method	Strains					
		precursor	c-atom correspondance		IBT100081		IBT100082		IBT100083	
					SFL	Dev	SFL	Dev	SFL	Dev
Glc	331	Glc	1,2,3,4,5,6	PAA	55.76	1.84	54.79	2.61	54.77	1.73
Gly	85	3PG	2	DMFDMA	1.12	0.14	1.13	0.05	1.16	0.06
Gly	144		1,2	DMFDMA	4.45	0.58	4.43	0.27	3.30	0.55
Gly	175		1,2	ECF	4.11	0.58	4.97	0.89	6.73	1.45
Ser	132		2,3	ECF	24.57	0.72	24.69	0.15	24.90	0.20
Ser	175		1,2	ECF	2.63	0.37	2.35	0.43	2.34	0.11
Ala	99	PYR	2,3	DMFDMA	25.17	0.33	25.31	0.30	25.39	0.42
Ala	116		2,3	ECF	25.61	0.18	25.98	0.30	26.06	0.31
Ala	158		1,2,3	DMFDMA	27.33	0.11	27.86	0.41	28.83	0.57
Val	127		2,2,3,3	DMFDMA	50.22	0.35	50.67	0.34	50.96	0.48
Val	143		1,2	DMFDMA	5.41	0.10	5.41	0.14	5.66	0.21
Val	144		2,2,3,3	ECF	50.53	0.31	51.39	0.44	51.69	0.37
Val	186		1,2,2,3,3	DMFDMA	53.54	0.49	54.86	0.65	55.36	0.91
Leu	158	PYR+AcCoA	2,2,3,3+2	ECF	72.61	0.31	73.39	0.69	73.96	0.28
Asp	115	OAA	2	DMFDMA	1.55	0.19	1.16	0.09	1.27	0.12
Asp	188		2,3,4	ECF	27.45	0.22	28.24	0.19	28.18	0.43
Asp	216		1,2,3,4	DMFDMA	29.40	0.66	29.81	0.39	27.50	2.25
Thr	146		2,3,4	ECF	25.70	1.06	26.22	0.27	26.49	0.21
Thr	175		1,2	ECF	2.26	0.15	2.58	0.11	2.54	0.08
Ile	158	OAA+PYR	2,3,4+2,3	ECF	53.51	0.34	54.11	0.33	54.36	0.48
Glu	143	2KG	1,2	DMFDMA	27.56	3.15	24.84	1.90	25.30	0.69
Glu	202		2,3,4,5	ECF	49.31	0.19	50.38	0.49	50.87	0.51
Glu	230		1,2,3,4,5	DMFDMA	52.17	0.63	52.11	0.47	52.64	0.56
Pro	142		2,3,4,5	ECF	49.87	0.26	50.83	0.36	51.20	0.50
Lys	156	2KG+AcCoA	2,3,4,5+2	ECF	72.70	0.28	74.04	0.51	74.36	0.29
Phe	143	PEP	1,2	DMFDMA	2.23	0.08	2.22	0.04	2.29	0.09
Phe	192	PEP+E4P	2,2,3,3+1,2,3,4	ECF	63.65	0.66	64.74	0.67	64.85	0.43

The consistency of the SFL data was checked for all cultivations. For example, fragments Ala99, Ala 116, Val127 and Val144 are derived from the second and third atom of pyruvate. If the measurements were consistent, the labeling of the fragment Val127 and Val144 should be as twice as the labeling of the fragment Ala99 and Ala116. The same calculation was applied to the fragment of Asp188 and Thr146 which are derived from the second to the fourth atoms of oxaloacetate. Both fragments show the same SFL value, indicating that the data obtained from the experiment were highly consistent.

6.3.8 Identification of pathways leading to pyruvate

From Table 6.4, it can be seen that the key metabolites such as 3-phosphoglycerate (3PG), pyruvate, oxaloacetate (OAA), acetyl-CoA and α-ketoglutarate (AKG) are the precursors for the different kind of anino acids. Therefore, considering these amino acid fragments labeled in the different positions, it is possible not only to assess the SFLs of some key metabolites in the central carbon metabolism of *S. cerevisiae*, such as the Embden-Meyerhof-Parnas (EMP), and tricarboxylic acid (TCA) cycle, but also to identify the activity of different pathways which is the first step in the metabolic network analysis.

When [1-^{13}C] glucose is metabolized into pyruvate via EMP pathway, 50% of the formed pyruvate is labeled in position 3. If pyruvate comes from Entner-Doudoroff (ED) pathway, 50% of pyruvate is labeled in position 1. When the pentose phosphate (PP) pathway is active, the labeled carbon atom in the first position of glucose will be lost in the first step as carbondioxide, resulting in no labeled carbon atom in pyruvate. The carbon atom labeled in the first position of pyruvate can be calculated by the subtraction of the average between Ala116 and Ala99 from Ala158. The subtraction of the percentage of carbon atoms labeled in the first position from Val142 yields the SFL of the second carbon atom of pyruvate which can be subtracted from the average SFL of Ala116 and Ala99, giving the percent of carbon atom of pyruvate labeled in the third position. The changes of pyruvate labeled in each position throughout the cultivation period for three different strains are visualized in Figure 6.7.

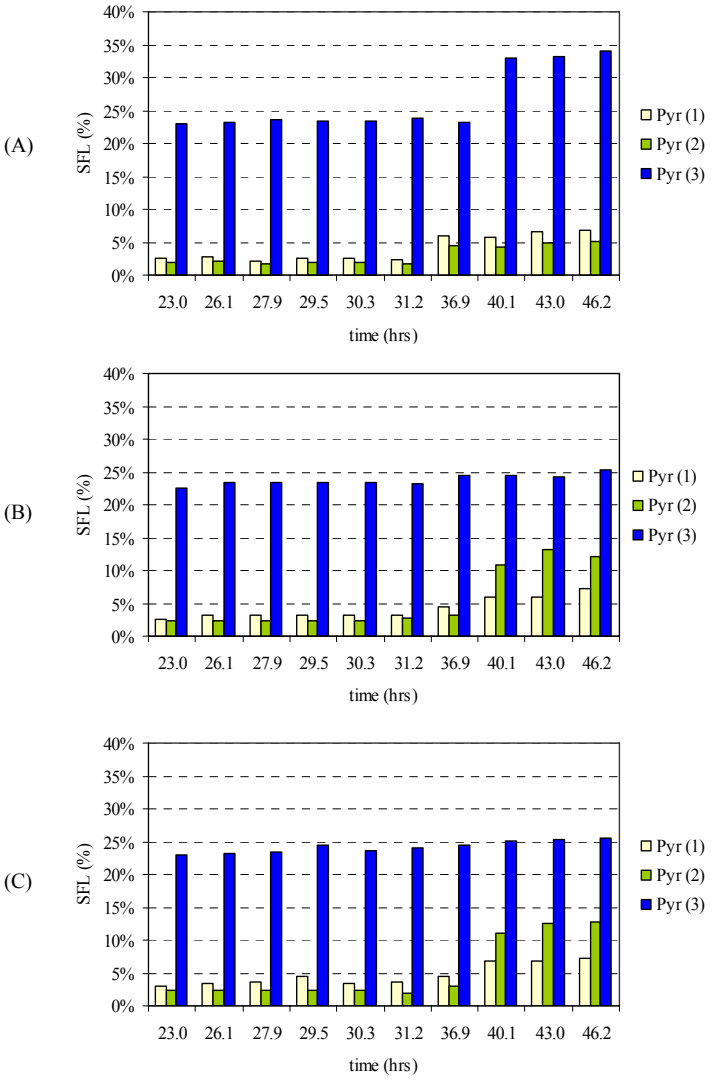


Figure 6.7. Change of pyruvate atom labeling at position 1, 2 and 3 during the aerobic cultivations of three different strains when 50% of [1-^{13}C]-labeled glucose was employed. (A) IBT100081, (B) IBT100082, and (C) IBT100083.

From Figure 6.7, it can be seen that the EMP pathway is the major pathway that is active for all strains as the SFL of the pyruvate labeled in the 3rd position are the highest. In the glucose phase where the first six peaks are, the SFL of pyruvate labeled in the 1st and 2nd positions are approximately between 2-3 % for every strain, which are a little bit more than natural labeling (1.1 %), indicating that ED pathway and the back flux from the TCA cycle are less active compared to the EMP pathway. In the ethanol phase, illustrated by the

last four peaks, it is seen that the SFL of pyruvate labeled in the 1st and 2nd position are increased, especially the 2nd position increased about four times. Interestingly, the increase of the pyruvate labeled in the 2nd position occurs only in the strains carrying the 6-MSAS gene, but not with the reference strain. This can possibly be explained by the back flux from the TCA cycle, for example malic enzyme converts the TCA cycle intermediate malic acid into pyruvate, which led to the introduction of the pyruvate labeled in the 2nd position, and it seems like this back flux is activated by the presence of the polyketide synthase gene. However, during the ethanol growth phase in the reference strain, the pyruvate labeled in the 3rd position increased to almost 35%, while those for the strain carrying 6-MSAS gene remained at 25% in the ethanol phase. An increased activity of malic enzyme during growth on ethanol is consistent with findings of dos Santos et al. (2003), who found that this flux increased during the consumption of acetate.

6.3.9 Metabolic flux calculation

The qualitative analysis of the SFL data above can provide useful information about the labeled position for each metabolite, leading to the indication of the active pathway in the organism. However, to quantify the metabolic fluxes in the central carbon metabolism of both the reference strain and the strains producing 6-MSA, computer simulation combined with mathematical modeling were used to estimate the fluxes. Only the data in the first exponential phase, where the glucose was used as a carbon source, were used in the calculation. The inputs for the model are the average of the SFLs in the first six samples, the flux of precursors to biomass for the respiro-fermentative batch cultivation (Gombert et al., 2001) and the yields of ethanol, acetate, and glycerol on glucose. The model assumes some independent fluxes and computes the remaining fluxes and also a set of SFL values for intracellular metabolites. The deviation between the measured SFL and the calculated SFL, as well as the difference between the measured fluxes and the calculated fluxes were generated during the simulation. The optimization routine minimizes the sum of these two deviations. The results of the best fit in the first exponential growth phase for three different strains which were normalized with respect to 100 mole of glucose consumption, are shown in Figure 6.8.

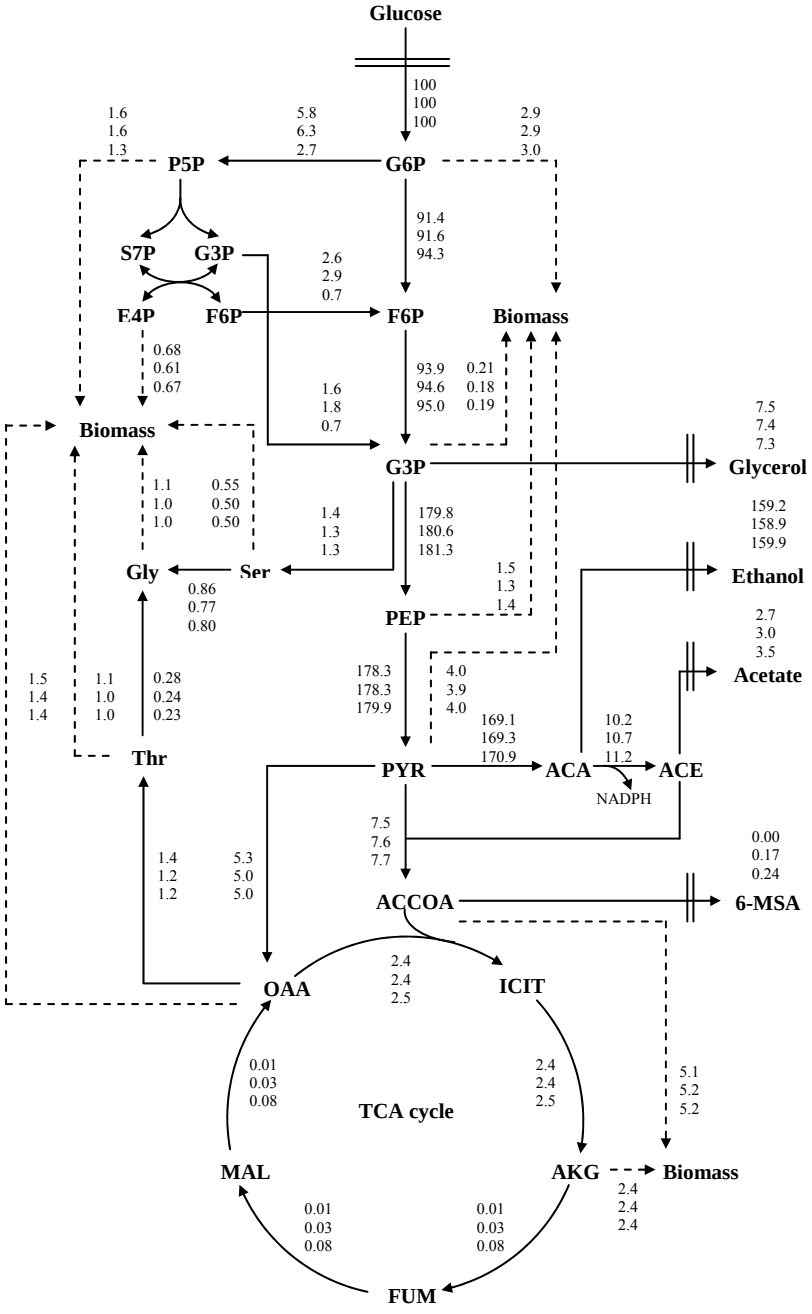


Figure 6.8. Flux distributions in the central metabolism of *S. cerevisiae* for three batch cultivations. Values in the first line correspond to the fluxes of the reference strain (IBT100081), whereas the values in the second and third line correspond to those of the strain IBT100082 and IBT100083, respectively. All the fluxes shown are the net fluxes, relative to the glucose uptake of 100 moles.

From Figure 6.8, the fluxes through each pathway for the three strains can be observed. The pathway that was mainly active for all the strains is the EMP pathway as approximately 90% of glucose came through this pathway. Among the three different strains, IBT100083 had the highest flux through the EMP pathway, followed by the strain IBT100082 and IBT100081, respectively. In the comparison between IBT100082 and IBT100081 (reference strain), the flux through the EMP pathway of IBT100082 was more than that of IBT100081. This is in accordance with the transcription data (Figure 6.4), which shows that the gene responsible for enzymes hexokinase (Hxk1), glucokinase (Glk1), phosphofructokinase (Pfk1), glyceroldehyde-3-phosphate dehydrogenase (Tdh1), phosphoglyceromutase (Gpm2), and enolase (Eno1) were included in subnetwork and all up-regulated in the strain IBT100082 compared to the reference strain. Furthermore, the gene encoding enzyme phophoglucoisomerase (Pgi1) responsible for the conversion of glucose-6-phosphate (G6P) to fructose-6-phosphate (F6P) which was not included in the subnetwork, was also up-regulated in IBT100082. The up-regulation of the genes and the increase in the flux through the EMP pathway in the strains producing 6-MSA, especially for the strain IBT100083, indicated that they needed to generate a high amount of acetyl-CoA which is used as precursor for 6-MSA production.

The PP pathway only carried a flux of about 2-6% of the incoming glucose flux, and the results show that the strain IBT100082 had a higher flux compared to IBT100081 (reference strain) and IBT100083. In the oxidative phase, IBT100082 gave 6.36 mol of pentose-5-phosphate (P5P)/100 mol of glucose, whereas the reference strain and IBT100083 gave 5.77 and 2.67 mol of P5P/100 mol of glucose, respectively (Figure 6.8). The increase of the flux through the PP pathway in the strain IBT100082 compared to the reference strain (IBT100081) is related to the up-regulation of the genes encoding enzymes G6P dehydrogenase (Zwf1), 6-phosphogluconolactonase (Sol4) which were included in the subnetwork, and enzyme 6-phosphogluconate dehydrogenase (Gnd1) which was not included in the subnetwork (Figure 6.4). The decrease of the flux through the PP pathway in the strain IBT100083 was probably because the flux from G6P was pushed towards the EMP pathway and therefore there was a lower flux through the PP pathway in this strain.

After pyruvate was generated by the EMP pathway, it was converted to acetyl-CoA (ACCOA) which subsequently entered the TCA cycle through reaction with oxaloacetate (OAA) to form citrate and further isocitrate (ICIT), and α-ketoglutarate (AKG), respectively (Figure 6.8). For the three strains the flux from acetyl-CoA and oxaloacetate to α-ketoglutarate was about the same. After the branching from α-ketoglutarate to biomass, the net flux leading to oxaloacetate is almost zero in all three strains.

When considering the flux from pyruvate through acetate (ACE), later converted to acetyl-CoA which is used as the precursor for 6-MSA production, it was found that the flux in this branch was increased in the strains carrying the 6-MSAS gene, especially the strain IBT100083 where *ACC1* is over-expressed. The flux from pyruvate to acetaldehyde (ACA) in the strains IBT100081, IBT100082 and IBT100083 were 169.07, 169.33 and 170.85 mol of acetaldehyde/100 mol glucose, and flux from acetaldehyde to acetate in those strains were 10.16, 10.67 and 11.17 mol of acetate/100 mol glucose, respectively.

Acetate produced by the cells can both be excreted out of the cells and converted by acetyl-CoA synthase (Acs1, 2) into acetyl-CoA. The total fluxes from acetate to acetyl-CoA and from pyruvate to acetyl-CoA in the strains IBT100081, IBT100082, and IBT100083 were 7.47, 7.64 and 7.71 mol of acetyl-CoA/100 mol glucose, respectively, indicating that there were more precursor supply available for the production of 6-MSA in the strain IBT100083 than other strains. In the comparison of IBT100082 and IBT100081, the flux calculation of this pathway corresponds to the expression data which shows that *THI3*, *PDB1*, *ALD3*, *ALD5*, *ACH1*, *ACS1*, and *ACS2* were up-regulated in the strain IBT100082 compared to IBT100081. However, for the strains IBT100082 and IBT100083 only 0.17 and 0.24 mol of 6-MSA/100 mol of glucose are required for 6-MSA biosynthesis and the changes of fluxes are therefore expected to be small, where it is interesting that there actually a transcriptional response.

In summary, the main changes in transcription profiles between the reference strain (IBT100081), the strain carrying 6-MSAS with its native promoter of *ACC1* gene (IBT100082) and the strain carrying 6-MSAS gene with over-expression of *ACC1*

(IBT100083) were observed in the batch cultivations. Amino acid and carbohydrate metabolism were substantially changed when comparing the strain IBT100081 and IBT100082 in the glucose and ethanol phase, respectively, as many amino acids and carbohydrate compounds were identified as reporter metabolites and some of the neighbouring enzymes around reporter metabolites were included in the subnetwork structure. Lipid metabolism was also found to be significantly influenced after the change of the *ACC1* promoter, which results in higher production of 6-MSA produced out of the cells. The metabolic network analysis was performed for quantitative inspection of metabolism. The result indicated that the fluxes in glycolysis, TCA cycle as well as fluxes from pyruvate through acetyl-CoA in the cytosol were increased in the strains carrying 6-MSAS gene. However, redirecting the fluxes leading to the secreted metabolites ethanol and acetate towards increased production of acetyl-CoA may result in an increased production of 6-MSA.

6.4 References

Christensen B, Nielsen J. 1999. Isotopomer analysis using GC-MS. Metab Eng. 1(4): 282-290.

Christensen B, Nielsen J. 2000. Metabolic network analysis. A powerful tool in metabolic engineering. Adv Biochem Eng Biotechnol. 66: 209-231.

dos Santos MM, Gombert AK, Christensen B, Olsson L, Nielsen J. 2003. Identification of in vivo enzyme activities in the co-metabolism of glucose and acetate by *Saccharomyces cerevisiae* using ^{13}C-labeled substrates. Eukaryotic Cell 2(3): 599-608.

Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. 2003. Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. Genome Res. 13(2): 244-253.

Gombert AK, dos Santos MM, Christensen B, Nielsen J. 2001. Network Identification and Flux Quantification in the Central Metabolism of *Saccharomyces cerevisiae* under Different Conditions of Glucose Repression. J. of Bacteriology 183(4): 1441-1451.

Grotkjær T, David H. 2005. C13 developed in connection with the PhD course "Linear and non-linear optimisation methods'' at Center for Microbial Biotechnology, BioCentrum-DTU, from June to December 2002.

Kealey JT, Liu L, Santi DV, Betlach MC, Barr PJ. 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proc. Natl. Acad. Sci. 95: 505-509.

Li C, Wong WH. 2001. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci USA. 98: 31-36.

Patil KR, Nielsen J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc Natl Acad Sci USA 102: 2685-2689.

Seker T, Møller K, Nielsen J. 2005. Analysis of acyl CoA ester intermediates of the mevalonate pathway in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 67: 119-224.

van den Berg MA, Jong-Gubbels P de, Kortland CJ, Dijken JP van, Pronk JT, Steensma HY. 1996. The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation. J. Biol. Chem. 271: 28953-28959.

van den Berg MA, Jong-Gubbels P de, Steensma HY. 1998. Transient mRNA responses in chemostat cultures as a method of defining putative regulatory elements: application to genes involved in *Saccharomyces cerevisiae* acetyl-coenzyme A metabolism. Yeast 14: 1089-1104.

Wattanachaisaereekul S, Lantz AE, Nilesen ML, Andrésson OS, Nielsen J. 2007. Optimization of heterologous production of the polyketide 6-MSA in *Saccharomyces cerevisia.* Biotechnol Bioeng. 97(4): 893-900.

Wiechert W, de Graaf AA. 1997. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol Bioeng. 55: 101-117.

Wittmann C, Heinzel E. 1999. Mass Spectrometry for metabolic flux analysis. Biotechnol. Bioeng. 62: 739-750.

Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielsen HB, Saxild HH, Nielsen C, Brunak S, Knudsen S. 2002. A new non-linear normalization method for reducing variability in DNA microarray experiments. Genome Biol. 3(9): research 00048.

Chapter 7

In Silico Metabolic Engineering for Improved Production of 6-MSA in *S. cerevisiae*

Abstract

In silico evaluation of gene deletion in *S. cerevisiae* carrying 6-MSAS gene from *P. patulum* was performed using genome-scale reconstructed metabolic model in order to identify the deletion target to increase the flux through 6-MSA. By using the minimization of metabolic adjustments (MOMA) approach and the OptKnock algorithm, *GDH1* was found as a gene deletion target with maximum yield of 0.04 mmol 6-MSA/g dw/hr at the μ_{max} of 0.23 hr^{-1}. Through the double deletion, the deletion of *GDH1* and *SER3* yields even higher 6-MSA at 0.07 mmol/g dw/hr with the μ_{max} of 0.22 hr^{-1}.

7.1 Introduction

Through metabolic engineering it is possible to introduce targeted genetic changes and engineer the metabolism of *S. cerevisiae* carrying a fungal PKS gene with the objective to obtain higher level production of 6-MSA at a high specific growth rate. However it is difficult to predict the effects of introducing genetic changes in yeast since its metabolism is often subjected to complex gene and enzyme regulations, and is constrained by mass and energy conservation laws on a numerous number of intracellular metabolites (Patil et al., 2005). A genome-scale model of *S. cerevisiae* (Förster et al., 2003) comprising the different levels of information, primarily on the stoichiometry of different reactions offer a suitable platform for developing the tools for analyzing and engineering metabolism (Patil et al., 2004). Regardless of the kinetics and regulation of the metabolism, it is possible to

at least partly predict the behaviour of the yeast metabolism by using analysis based on genome-scale stoichiometric models.

7.2 Materials and methods

7.2.1 Metabolic model

Genome scale reconstruction of *S. cerevisiae* suggested by Förster et al. (2003) was used as stoichiometric model of yeast metabolism. The simulations were performed under the aerobic glucose-limited conditions. The glucose uptake rate was fixed to 3 mmol/g dw/hr and the oxygen uptake rate was kept constantly of 9 mmol/g dw/hr.

7.2.2 Flux balance analysis (FBA), minimization of metabolic adjustments (MOMA) and genetic algorithm

FBA, MOMA and genetic algorithm were obtained from Patil et al. (2005)

7.3 Results and discussion

Genome-scale stoichiometric models represent the integrated metabolic potential of a micro-organism. Through flux-balance analysis where constraints on fluxes are imposed through balances on all intracellular metabolites it is possible to identify all feasible metabolic phenotypes such as the fluxes through all metabolic reactions under steady state conditions (Förster et al., 2003). The approaches to determine the metabolic phenotypes are to use flux balance analysis (FBA) and minimization of metabolic adjustments (MOMA). These approaches were extensively reviewed in Kauffman et al. (2003) and Segre et al. (2002). Using the genome-scale metabolic model, FBA was applied to evaluate the potential for production of 6-MSA in yeast. The reactions towards the polyketide were added into the model and both the flux towards 6-MSA production and the specific growth rate were optimized using linear programming. Hereby the theoretical yields of 6-MSA (mmol/g dw/hr) at different specific growth rates were calculated and the results are shown in Figure 7.1.

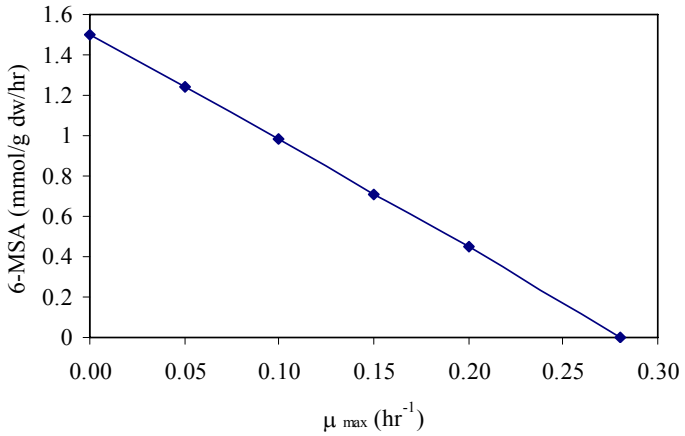


Figure 7.1. The theoretical yield of 6-MSA from *S. cerevisiae* at different growth rates suggested by the genome scale model using an FBA approach.

In order to identify possible targets for metabolic engineering to improve the production of 6-MSA a gene deletion analysis was performed. Each gene encoding enzymes participating in the metabolism was deleted and the 6-MSA yield was calculated. The gene deletions obtained using the MOMA approach and the OptKnock algorithm for improving the yield of 6-MSA are summarized in Table 7.1.

Table 7.1. Different gene deletions suggested by MOMA approach for improving the yield of 6-MSA in *S. cerevisiae*.

Number of deletions	Suggested gene deletions*	6-MSA yield (mmol/g dw/hr)	μ_{max} (hr^{-1})
1	*GDH1*	0.040	0.23
2	*GDH1, SER3,*	0.070	0.22
	GDH1, DAL7	0.061	0.24
	GDH1, ICL1	0.059	0.24
	GDH1, MTD1	0.051	0.23
	GDH1, THR1	0.045	0.24
	GDH1, YHL012W	0.042	0.24
	GDH1, FDH2	0.042	0.24
	GDH1, PRO2	0.041	0.24
	GDH1, PRO1	0.041	0.24
	GDH3, PCK1	0.041	0.24

* *GDH1* : NADPH(+)-dependent glutamate dehydrogenase
SER3 : 3-Phosphoglycerate dehydrogenase
DAL7 : Malate synthase
ICL1 : Isocitrate lyase
MTD1 : NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase
THR1 : Homoserine kinase
YHL012W : Putative protein of unknow function, has some homolgy to Ugp1p which encodes UDP-glucose pyrophosphorylase
FDH2 : NAD(+)-dependent formate dehydrogenase
PRO2 : Gamma-glutamyl phosphate reductase
PRO1 : Gamma-glutamyl kinase
PCK1 : Phosphoenolpyruvate carboxykinase

Reported productivities of 6-MSA are about 0.022 mmol 6-MSA/g dw/hr at a μ_{max} of 0.26 hr^{-1}, and in a strain with overexpression of *ACC1* it is about 0.031 mmol 6-MSA/g dw/hr (μ_{max} of 0.27 hr^{-1}). The comparison of these 6-MSA productivities to those in Table 7.1 indicates that through deletion of the target genes, the productivities of 6-MSA can be substantially increased due to the increase of NADPH and acetyl-CoA which are used as precursors for 6-MSA biosynthesis, especially by deletion of *GDH1* and *SER3* resulting in 6-MSA specific productivity of about 0.07 mmol/g dw/hr. Thus the *in silico* gene deletion analysis could eventually be a tool and platform for designing an efficient yeast platform for the production of polyketides.

7.4 References

Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. 2003. Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. Genome Res. 13(2): 244-253.

Kauffman KJ, Prakash P, Edwards JS. 2003. Advances in flux balance analysis. Curr. Opin. Biotechnol. 14: 491-196.

Patil KR, Åkesson M, Nielsen J. 2004. Use of genome-scale microbial models for metabolic engineering. Curr. Opin. Biotechnol. 15: 64-69.

Patil KR, Rocha I, Förster J, Nielsen J. 2005. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6: 308.

Segre D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic network. Proc Natl. Acad. Sci. 99: 15112-15117.

Chapter 8

Cloning, Characterization and Heterologous Expression of *atX* from *A. terreus* IBT12713 and the Hybrid 6-MSA Synthase Gene from *P. patulum* and *A. terreus* in *S. cerevisiae*

Abstract

Since the 6-MSAS gene from *Penicillium patulum* has been successfully expressed in *S. cerevisiae* and high production of 6-MSA was obtained, it is interesting to clone the 6-MSAS gene from another organism and expressed this in yeast to investigate whether the production of 6-MSA is affected by the source of the gene. *Aspergillus terreus* has been reported to hold the gene *atX* which encodes a 6-MSAS. Here, *atX* from *A. terreus* IBT12713 was amplified by RT-PCR. The amino acid sequence of AtX protein in this study was characterized and found to have 99% identity to the AtX from other studies, and it showed 63% identity to the 6-MSAS from *P. patulum*. The *atX* from *A. terreus* IBT12713 was transformed into *S. cerevisiae* with co-expression of *npgA* from *A. nidulans*, but no 6-MSA was produced. The *atX* was modified by changing the 5′ terminal to that of 6-MSAS gene from *P. patulum*, resulting in a hybrid 6-MSAS gene which contained the gene from *P. patulum* at 5′ terminal and the gene from *A. terreus* at 3′ terminal. The hybrid 6-MSAS gene was co-transformed with the *npgA* from *A. nidulans* in *S. cerevisiae*, but again no 6-MSA could be detected. An explanation for the unsuccessful heterologous production of 6-MSA from *A. terreus* in *S. cerevisiae* might be due to codon bias and the formation of secondary structures of mRNA which could inhibit the translation process of mRNA into protein. An engineered *atX* was proposed by changing

the rare codons into the codons that are abundant in yeast and this should also prevent formation of secondary mRNA structures. However, the adapted *atX* sequence was not experimentally evaluated.

8.1 Introduction

Fungi are eukaryotic organisms that are more related to animals than plants. As they are heterotrophic, fungi share with animals the ability to export the hydrolytic enzymes to break down organic materials which can be absorbed for nutrition and continue the cycle of nutrients through the ecosystems (White and Arora, 2004). Many fungal species are free-living saprobes in woody substrates, soils, leaf litter, dead animal and animal exudates. However, a large diversity of other fungi are biotrophs or living symbiotic in associations with plants, algae, animals and prokaryotes such as lichens (Ahmadjian, 1993), mycorrhizae, and leaf or stem endophytes (Smith and Read, 1997).

Although fungi can cause immense economic losses due to their harmful activities as saprotrophs, including the damage to food and manufactured products as well as the parasitic activities that cause heavy crop losses (Stokstad, 2007) and diseases of humans and domestic animals (Rippon, 1988), they still provide a rich source of enzymes and metabolites that are of great economic importance. They have been long exploited as food and in food processing (Niederhaus and Stahl, 2003). Since the fermentation industry developed, fungi been used to produce an increasing range of valuable products including the polyketide metabolites which can be used as antibiotics and other drugs, e.g. antitumor and anticholesterolemic properties.

So far only a few fungal polyketides have been studied in detail in order to understand their biosynthesis pathways and regulations. Those studies include fermentation optimization and heterologous expression of polyketide synthases in surrogate hosts. Heterologous expression of polyketide synthase genes is emerging as a viable alternative to both classical strain and fermentation process development, and molecular biological manipulation of the native producer strain (Pfeifer and Khosla, 2001). Heterologous

expression of fungal polyketide can provide strains with better fermentation characteristic that are more amenable to further optimization, or present the fermentation alternative for the symbiont-derived metabolites.

Amongst the polyketides produced by fungi, 6-MSA is considered as a model fungal polyketide from *P. patulum* as it is the most well-characterized fungal polyketide. The gene encoding the enzyme responsible for the production of 6-MSA from *P. patulum* was heterologous expressed and intensively studied in *S. cerevisiae*. However, 6-MSA is naturally produced not only by *P. patulum*, but it is produced by *P. expansum* (White et al., 2006), *A. terreus* (Fujii et al., 1996), and *Byssochlamys nivea* (Puel et al., unpublished) as well.

The objective of this study is to investigate whether the production of 6-MSA in *S. cerevisiae* is affected by the source of the gene. Therefore, the 6-MSAS gene from *A. terreus* was cloned and heterologously expressed in *S. cerevisiae* in order to compare the production of 6-MSA to that from *P. patulum*.

8.2 Materials and methods

8.2.1 Plasmids and strains

The plasmids and strains used in this study are listed in Table 8.1.

Table 8.1. List of plasmids and strains

	Characteristics	Reference
Plasmids		
pRS426CT	Expression vector containing *TEF1* promoter, *CYC1* terminator, and *URA3* marker	Wattanachaisaereekul et al. (submitted)
pRS424CTnpgA	Expression vector containing *TEF1* promoter, *CYC1* terminator, *TRP1*marker, and *npgA* from *A. nidulans*	Wattanachaisaereekul et al. (submitted)
pRS426CTMSA-AT	Expression vector containing *TEF1* promoter, *CYC1* terminator, *URA3* marker, and 6-MSAS gene from *A. nidulans*	This study
pRS426CTMSA-PP-AT	Expression vector containing *TEF1* promoter, *CYC1* terminator, *URA3* marker, and hybrid 6-MSAS gene from *P. patulum* and *A. nidulans*	This study
Strains		
A. terreus IBT12713	Wild type of *A. terrues* producing 6-MSA	Our laboratory
S. cerevisiae CEN.PK 113-9D	MATa;*ura3-52;HIS3;LEU2;trp1-289;MAL2-8;SUC2*	Peter Kötter [a]
S. cerevisiae IBT100085	CEN.PK 113-9D with the change of *ACC1* promoter to *TEF1* promoter, and carrying plasmids pRS426CTMSA-AT and pRS424CTnpgA	This study
S. cerevisiae IBT100086	CEN.PK 113-9D with the change of *ACC1* promoter to *TEF1* promoter, and carrying plasmids pRS426CTMSA-PP-AT and pRS424CTnpgA	This study

[a] Institut für Mikrobiologie, der Johann Wolfgang Goethe-Universität Frankfurt, Germany

8.2.2 Fungal cultivation and total RNA extraction

The spore suspension of *A. terrues* IBT12713 was cultivated in 100 mL minimal media in 500 mL shake flasks. The medium consisted of 20% w/v glucose 50 mL/L, 1M $NaNO_3$ 2 mL/L, and mineral mix 4 mL/L. Mineral mix was prepared by adding KCl 26 g, $MgSO_4.7H_2O$ 26 g, KH_2PO_4 76 g and trace element solution 50 mL in distilled water with the final volume of 1L. The trace element solution for 1 L consisted of $CuSO_4.5H_2O$ 0.4 g, $Na_2B_4O_7.10H_2O$ 0.04 g, $FeSO_4.7H_2O$ 0.8 g, $MnSO_4.2H_2O$ 0.8 g, $Na_2MoO_4.2H_2O$ 0.8 g, and $ZnSO_4.7H_2O$ 8.0 g. About 36 hr cultivation, the fungal mycelia were harvested and

filtrated through the mira-cloth. The medium absorbed by the mycelia was removed by pressing the mycelia inside the mira-cloth until they were dried. The mycelia were instantaneously ground into a fine powder in a mortar under the liquid nitrogen. The powder was then processed for the RNA isolation using NucleoSpin® RNA Plant Kit (Macherey-Nagel GmBH & Co. KG, Düren, Germany). To concentrate the extracted RNA, it was precipitated from the aqueous phase, followed by the addition of 0.25 volume of isopropanol and 0.25 volume of RNA precipitation solution (1.2 M NaCl, 0.8 M $C_6H_5Na_3O_7.2H_2O$). After thorough mixing, store the final solution for 10 minutes at room temperature. The precipitated RNA was collected by centrifugation at maximum speed for 10 minutes at 4 °C in a microfuge. The pellet was washed twice with 75% ethanol and centrifuged again. Any remaining ethanol was removed with a disposable pipette tip. The open tube was stored on the bench for a few minutes to allow the ethanol to evaporate and then 50 µL of DEPC-treated water was added and the RNA solution was stored at -80 °C until further processing.

8.2.3 Reverse transcription and polymerase chain reaction (RT-PCR) and yeast transformation

The expand reverse transcriptase from *E. coli* AP401 (Basel, Switzerland) was used for reverse transcription reaction. Oligonucleotides were synthesized by MWG-Biotech, AG, Germany, are listed in Table 8.2. PCR reactions were performed using Phusion polymerase according to the manufacturer's guideline. GFX spin columns from Amersham Pharmacia were used for purification of DNA from both gel band and solution. After the purification, the whole 6-MSAS gene was sent to MWG-Biotech, AG, Germany, for sequencing. Transformation of yeast cells was carried out according to Gietz and Woods (2002). *E. coli* DH5α was used for propagation of recombinant plasmids. The preparation of competent *E. coli* and transformation protocol were performed according to Sambrook and Russell (2001).

Table 8.2. The primers used for the amplification of 6-MSAS gene from *A. terreus*. The bold represents 6xHis Tag, and the underline sections indicate the restriction sites introduced to facilitate cloning.

Primers	Sequence (5´- 3´)
	*Hind*III *Spe*I *Bam*HI *Pst*I *Pac*I *Kpn*I
SWA23	CCCAAGCTTTAGAACTAGTGGATCCCTGCAGGTTAATTAAGGAGGTACCTTTAAAATGGAGGTACATGGAGATGAAGTG
	*Eco*RI *Xho*I *Asc*I *Swa*I 6xHis Tag
SWA24	CCGGAATTCTACATGACTCGAGGGCGCGCCATTTAAATTA**ATGATGATGATGATGATG**ATTTCCCATCTTTTCCAAAAA
	*Bcl*I
SWA25	AAGTATTGATCATACAAGGCCGGACGGACCTC
SWA26	GGGACCTAGACTTCAGGTTGTC

8.2.4 Yeast cultivation and analysis of 6-MSA

In order to check the production of 6-MSA, the yeast carrying the plasmids with the 6-MSAS gene from *A. terreus* and the *npgA* from *A. nidulans* was cultivated in the 100 mL minimal media medium in 500 mL baffled shake flasks. The medium consisted of glucose 10 g/L, $(NH_4)_2SO_4$ 7.5 g/L, KH_2PO_4 14.4 g/L, $MgSO_4.7H_2O$ 0.5 g/L, 0.05 mL/L Antifoam 298 (Sigma-Aldrich, St. Louis, MO, USA) and 2 mL/L trace metal solution ($FeSO_4.7H_2O$ 3 g/L, $ZnSO_4.7H_2O$ 4.5 g/L, $CaCl_2.6H_2O$ 4.5 g/L, $MnCl_2.2H_2O$ 0.84 g/L, $CoCl_2.6H_2O$ 0.3 g/L, $CuSO_4.5H_2O$ 0.3 g/L, $Na_2MoO_4.2H_2O$ 0.4 g/L, H_3BO_3 1 g/L, KI 0.1 g/L, and $Na_2EDTA.2H_2O$ 15 g/L), and 2 mL/L vitamin solution (d-biotin 50 mg/L, Ca-pantothenate 1 g/L, thiamin-HCl 1 g/L, pyridoxin-HCl 1 g/L, nicotinic acid 1 g/L, p-aminobenzoic acid 0.2 g/L, and m-inositol 12.5 g/L). pH of the medium was adjusted to 6.5 by 2M NaOH prior to autoclaving. The cultures were incubated at 30 ^oC, with shaking at 150 rpm (model 3033, GFL, Burgwedel, Germany).

After 48 hours fermentation, the fermentation broth was filtered through a 0.45 μm pore-sized cellulose acetate filter (GE Osmonics Labstore, Minnetonka, MN). The filtrate was quantified for 6-MSA by HPLC (Agilent 1100 series) with Luna C18(2) column, 150 × 4.60 mm with 5 micron of the porous silica particles to which the C18 phase is bonded, using a gradient of 50 ppm trifluoroacetic acid (TFA) in milliQ water (solvent A) and 50

ppm TFA in acetonitrile (solvent B) at a flow rate of 1 mL/min. The gradient of the solvents was 20% to 60% B in 10 minutes, then 20% B for the next 2 minutes.

8.3 Results

8.3.1 Construction of the plasmid and strain

The purified RNA from *A. terrues* was used in the process of making cDNA of the 6-MSAS gene by the reverse transcription reaction by using primer SWA24 which contain the 6xHis tag as a sequence specific reverse primer. In the PCR reaction primers SWA23 and SWA24 (Table 8.2) were used for amplification of the 6-MSAS gene. To obtain the plasmid pRS426CTMSA-AT, the PCR product which was carrying the *atX* encoding the 6-MSAS from *A. terreus* and amplified by RT-PCR was double digested by *Spe*I and *Eco*RI, and ligated into the *Spe*I/*Eco*RI site of plasmid pRS426CT. The plasmid pRS426CTMSA-AT was co-transformed with plasmid pRS424CTnpgA in *S. cerevisiae* CEN.PK113-9D to yield the *S. cerevisiae* IBT100085.

8.3.2 Characterization of *atX* and heterologous production in *S. cerevisiae*

The nucleotide sequence of *atX* was determined over the open reading frame as 5,430 bp (including the 6xHis Tag and stop codon) which was transcribed as 1,809 amino acids. (Appendix B.1). The open reading frame did not contain the intron as it was amplified from RNA. The sequence of *atX* from this study was compared with other 6-MSASs gene in order to investigate the similarity among the genes, and the results are summarized in Table 8.3. This indicates that *atX* from this study had a high similarity to the *atX* from other studies with 99% identity at both the nucleotide and the amino acid level, but it showed around 60% identity in the amino acid sequence to the 6-MSAS gene from *P. patulum, A. clavatus and Byssochlamys nivea*. *S. cerevisiae* IBT100085 carrying the plasmid with *atX* gene from *A. terreus* IBT12713 which was co-transformed with a plasmid bearing the *npgA* from *A. nidulans* (pRS424CTnpgA) was cultivated in glucose minimal media in shake flasks. Unfortunately no production of 6-MSA could be detected.

Table 8.3. Percent identity of the 6-MSAS gene from different organisms to the *atX* from *A. terreus* IBT12713 cloned by RT-PCR in this study

6-MSAS gene from different organisms	Similarity to *atX* from *A. tereues* IBT12713				Accession number
	Amino acid level		Nucleotide level		
	% identity	% coverage	% identity	% coverage	
A. terreus NIH2624	99.6	100	99.8	100	XM_00121543
A. terreus	99.4	100	99.2	100	U31329
A. terreus	98.4	100	98.1	100	D85860
A. clavatus	63.8	96.2	84.0	4.3	XM_001273092
B. nivea	63.3	99.3	82.4	3.4	AF360398
P. patulum	62.5	99.8	89.9	3.1	X55776

8.3.3 Construction of the hybrid 6-MSAS gene from *P. patulum-A. terreus* and heterologous production in *S. cerevisiae*

Since no 6-MSA was produced from *S. cerevisiae* carrying the *A. terreus atX*. The change of the 5′ region of some genes has shown to be absolutely necessary and apparently sufficient to obtain a good level of protein expression in different heterologous hosts (Batard et al., 2000). The front part of *atX* from *A. terreus* was therefore replaced with the front part of the 6-MSAS gene from *P. patulum*. The construction of the hybrid gene was firstly performed by searching for unique restriction sites that can easily be cut for swapping the DNA fragment. *Bgl*II which was located in the un-conserved region inside the acyltransferase (AT) domain of *P. patulum* 6-MSAS gene was used as the restriction site. *Bgl*II was not found in *atX* but it can be ligated with *Bcl*I. Therefore, a *Bcl*I restriction site was introduced in the primer SWA25. Two nucleotides ‘‘TA’’ were also included in the primer SWA25 at position 13 and 14 from 5′ end in order not to change the open reading frame after the ligation. However, the last part of *atX* was amplified by primer SWA25 and primer SWA26 by using plasmid pRS426CTMSA-AT as a template. The PCR product was cut by *Bcl*I/*Hind*III and ligated into the plasmid pRS426CTMSA-PP which had been cut by *Bgl*II and *Hind*III to yield the plasmid with hybrid 6-MSAS gene from *P. patulum* and *A. terreus*, pRS426CTMSA-PP-AT (Figure 8.1). To construct *S. cerevisiae* IBT100086, the plasmid pRS426CTMSA-PP-AT was co-transformed with

plasmid pRS424CTnpgA in yeast with the over-expression of *ACC1* to ensure the high level of malonyl-CoA. Twenty-five transformants were selected and cultivated in glucose minimal media in the shake flasks for 48 hours. 6-MSA could not be detected from any of them.

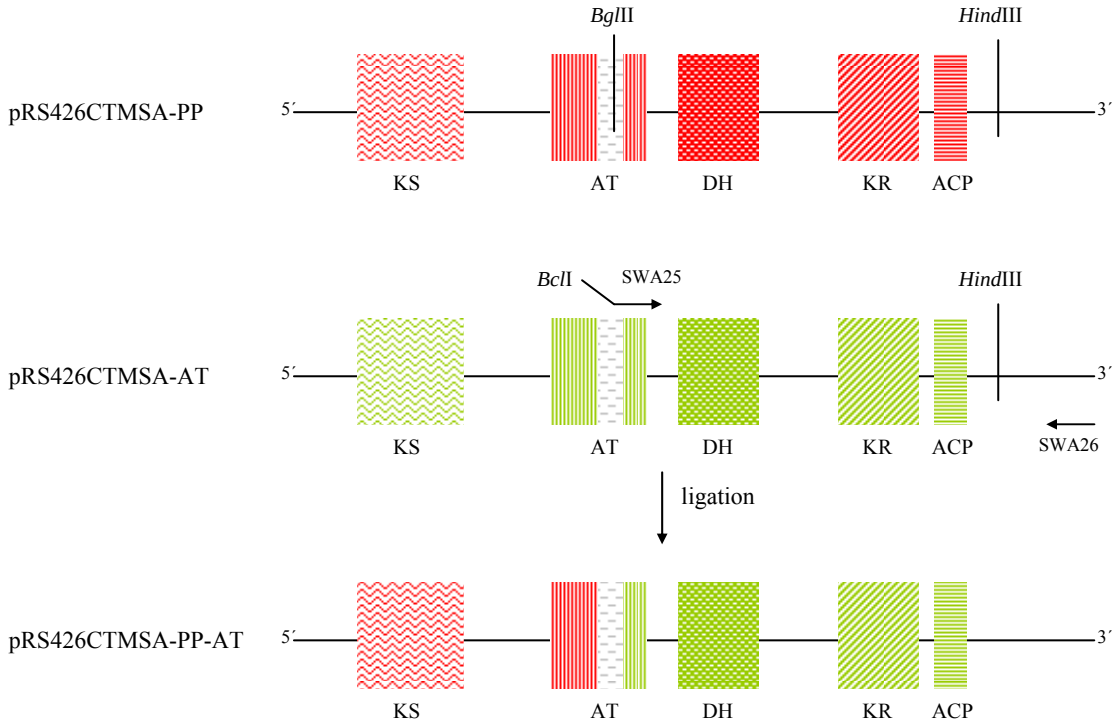


Figure 8.1. Construction of thy hybrid plasmid pRS426CTMSA-PP-AT which the 6-MSAS gene contains both parts from *P. patulum* (red) and *A. terreus* (green).

8.4 Discussion

There are several factors that play an important role in the expression of a gene in a heterologous host, such as the expression vector and the transcriptional promoter, the sequence around the N-terminal region, and the codon bias (Guftasson et al., 2004). Since the expression vector with the promoter used in this study was exactly the same as it was used for the expression of 6-MSAS gene from *P. patulum*, and since the change of the *atX* at the N-terminal to that from *P. patulum* did not lead to 6-MSA production, the main reason for lack of production of 6-MSA is believe to be codon bias. As the *atX* was

originally from *A. terreus*, there are some codons that are rarely used in *S. cerevisiae*, meaning that the tRNAs for these rare codons in yeast are scarce, thus the codon bias affects directly the translation process and the expression of the 6-MSAS protein. The frequency of codon usage for the whole genome of *S. cerevisiae* is shown in Appendix C.

Comparing the nucleotide sequence of *atX* from *A. terreus* with the codon usage of *S. cerevisiae*, we found that there are 70 sites in *atX* that are rarely used in *S. cerevisiae* (47 sites for rare codon with relative adaptiveness between 10-20, and 23 places for extremely rare codon with relative adaptiveness lower than 10). The 70 sites of codons in *A. terreus atX* which are rarely used in yeast are illustrated in Figure 8.2. Morover, by using GeneBee-Molecular Biology Server (Russia), we found that the mRNA transcribed from the *A. terrues atX* has a secondary structure which is not found in the mRNA of 6-MSAS gene from *P. patulum*. The secondary structure of the mRNA from *atX* was located at nucleotide position 4,628 to 4,882 (Figure 8.2) with a Gibbs free energy of -26.9 kcal/mol. The secondary structure with strong Gibbs free energy found here could inhibit the translation of mRNA into protein. By using Java Codon Adpatation Tool, Germany, the *atX* was *in silico* engineered such that the bias codons in *atX* were removed and replaced by the codons more frequently used in *S. cerevisiae*. The adapted *atX* sequence for the heterologous expression in yeast was shown in Appendix B.2. After the *atX* was adapted, the secondary structure of mRNA with strong energy between nucleotides was no longer found, indicating that the mRNA has a higher chance to be translated into protein. The structures of mRNA at position 4,628 to 4,882 before and after engineering are shown in Figure 8.3. However, the codon usage in *atX* from *A. terrues* and 6-MSAS gene from *P. patulum* which was successfully expressed in *S. cerevisiae* were compared in Table 8.4.

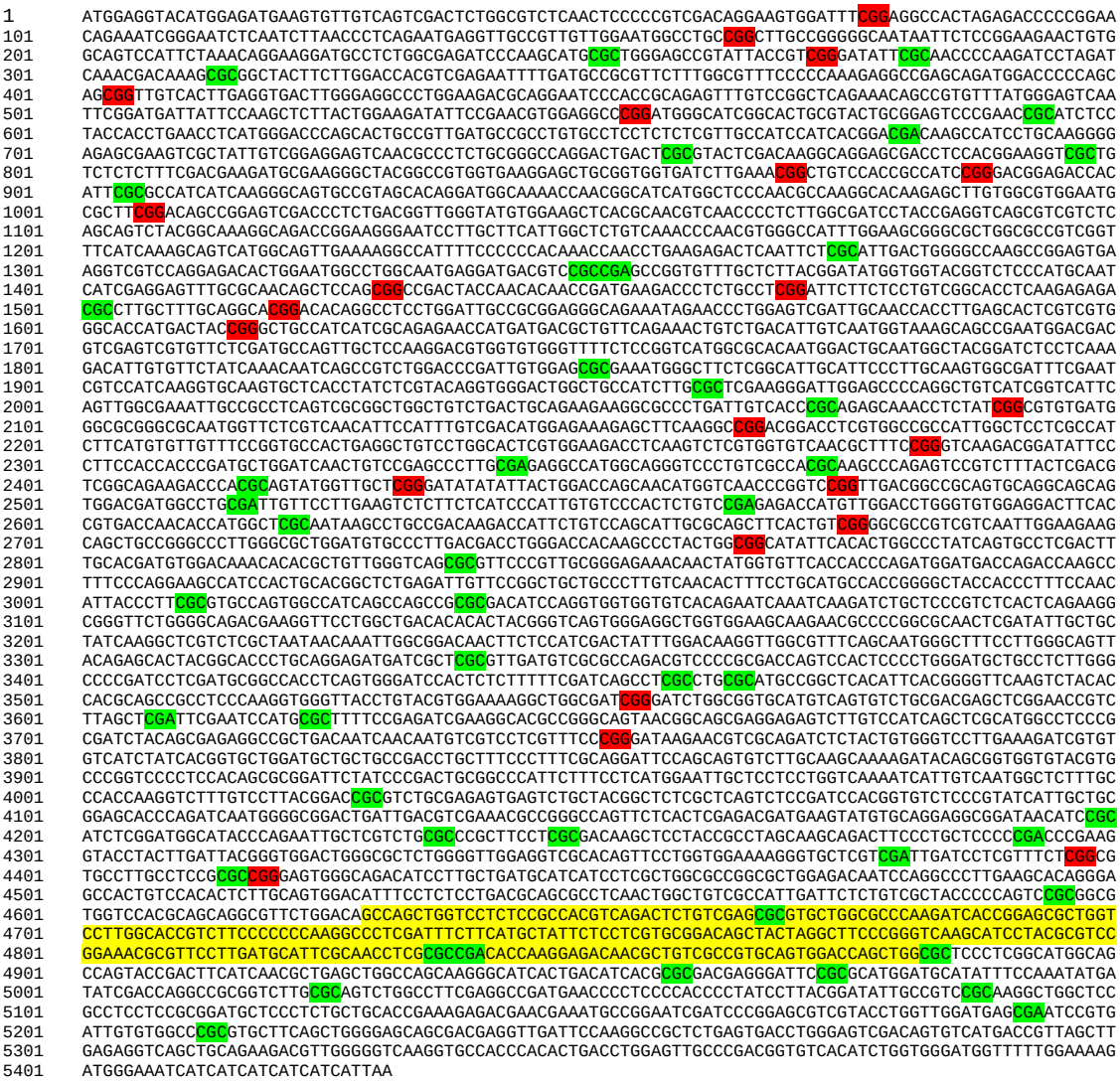
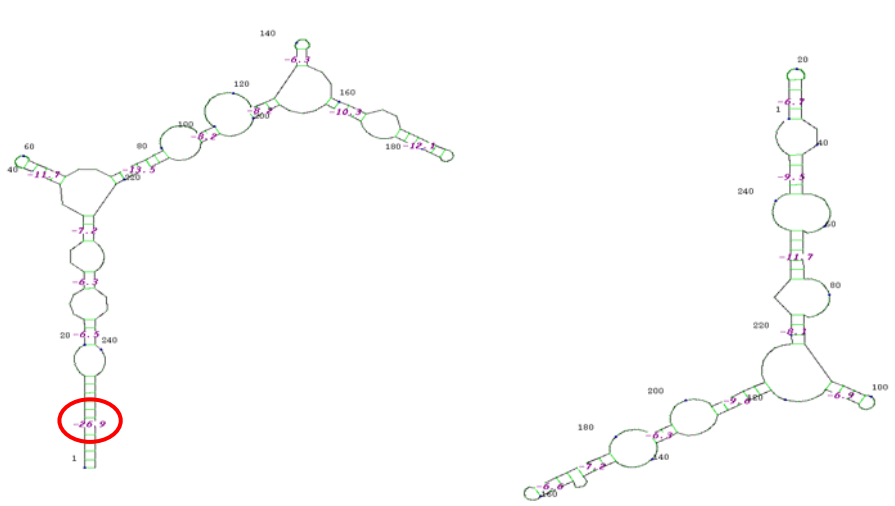


Figure 8.2. The distribution of codons in *atX* encoding 6-MSAS from *A. terreus* IBT12713 which are rarely used (relative adaptiveness 10-20; highlight in green), and are extremely rarely used in *S. cerevisiae* (relative adaptiveness lower than 10; hightlight in red). The highlight in yellow represents the area that the secondary structure with high energy has been found in mRNA.



A (before) B (after)

Figure 8.3. The structure of mRNA of *atX* at position 4,628 to 4,882 before and after *in silico* engineered the sequence. The large Gibbs free energy (-26.9 kcal/mol) between nucleotides indicated by the circle was found only in the structure of mRNA before reengineering.

Table 8.4. The codon usage in 6-MSAS gene from *P. patulum* compared with atX from *A. terreus* IBT12713 before and after engineering the gene.

Characteristics	6-MSAS gene from *P. patulum*	*atX* from *A. terreus* IBT12713	
		Before engineering	After engineering*
Codons that rarely found in yeast	50 places	47 places	0 place
Codons that extremely rarely found in yeast	18 places	23 places	0 place
mRNA Secondary structure with high energy	0 place	1 place	0 place
Successfully heterologous produced in yeast	yes	no	N/A

* The engineered *atX* has not been constructed yet. The sequence is shown in Appendix B.2.

In conclusion, we constructed the plasmid with *atX* encoding 6-MSAS gene from *A. terreus* IBT12713. Another plasmid was constructed by a 6-MSAS gene fusion with the first part on the gene from *P. patulum* and the last part from *A. terreus*. These plasmids were transformed in *S. cerevisiae* with co-expression of the plasmid carrying *npgA* from *A. nidulans*. No production of 6-MSA could be detected from both strains. This might be due to the codons in *atX* that are rarely used in *S. cerevisiae*. The sequence of *atX* was adapted in order to remove the codon bias and unfavorable mRNA secondary structure. However,

the adapted sequence need to be tested experimentally to investigate whether it can lead to the production of 6-MSA.

8.5 References

Ahmadjian V. 1993. The lichen symbiosis. John Wiley & Sons, New York. 266 p.

Batard Y, Hehn A, Nedelkina S, Schalk M, Pallett K, Schaller H, Werck-Reichhart D. 2000. Increasing expression of P450 and P450-reductase proteins from monocots in heterologous systems. Arch. Biochem. Biophys. 379(1): 161-169.

Fujii I, Ono Y, Tada H, Gomi K, Ebizuka Y, Sankawa U. 1996. Cloning of the polyketide synthase gene *atX* from *Aspergillus terreus* and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. Mol. Gen. Genet. 253(1): 1-10.

Gietz RD, Woods RA. 2002. Transformation of yeast by the LiAc/SS carrier DNA/PEG method. Methods in Enzymology 350: 87-96.

Gustafsson C, Govindarajan S, Minshull J. 2004. Codon bias and heterologous protein expression. Trends Biotechnol. 22(7): 346-353.

Niederhaus A, Stahl U. 2003. Fermented food production using genetically modified yeast and filamentous fungi. *In* Heller KJ (ed.) Genetically engineered food: Methods and dectection. Wiley-VCH, Germany. 287 p.

Pfeifer BA, Khosla C. 2001. Biosynthesis of polyketides in heterologous hosts. Microbiol. Mol. Biol. Rev. 65(1): 106-118.

Rippon JW. 1988. Medical mycology: The pathogenic fungi and the pathogenic actinomycetes. W. B. Saunders Company Harcourt Brace Jovanovich, Inc., Philadelphia, PA, 797 p.

Sambrook J, Russell DW. 2001. A laboratory manual: Protocol 24 The Inoue method for preparation and transformation of competent *E. coli*: "Ultra-Competent" Cells. Molecular Cloning (3rd editon) 1: 1.112-1.115.

Smith SE, Read DJ. 1997. Mycorrhizal symbiosis (2nd edition). Academic Press, San Diego. 605 p.

Stokstad E. 2007. Deadly wheat fungus threatens world's breadbaskets. Science 315: 1786-1787.

Wattanachaisaereekul S, Lantz AE, Nielsen ML, Nielsen J. 2007. Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. (Submitted).

White NA, Arora DK. 2004. The important of wood-decay fungi in forest ecosystems. Marcel Dekker Inc., UK.

White S, O´ Callaghan J, Dobson ADW. 2006. Cloning and molecular characterization of *Penicillium expansum* genes upregulated under conditions permissive for patulin biosynthesis. FEMS Microbiol. Lett. 255 (1): 17-26.

Conclusions and future perspectives

The overall objective of this dissertation was to study the heterologous expression of a fungal polyketide synthase in *S. cerevisiae*. 6-MSA, which is naturally produced by *P. patulum* was chosen as a model polyketide. Functional expression of the 6-MSAS gene in *S. cerevisiae* requires the co-expression of a heterologous phosphopantetheinyl transferase (PPTase) to convert the apo-PKS to its active holo form. PPTase genes from *B. subtilis* and *A. nidulans*, respectively, were used in this study. Using the *GAL1* promoter to drive the expression, a plasmid carrying the 6-MSAS gene from *P. patulum* and a plasmid carrying either one of the two PPTase genes were co-transformed into *S. cerevisiae*. The strains were characterized in batch cultivations with a galactose minimal media, and better productivity of the heterologous produced 6-MSA was obtained by the co-expression of PPTase from *A. nidulans*.

In order to improve the production of 6-MSA by enhancing the supply of precursors, the promoter of the gene (*ACC1*) encoding acetyl-CoA carboxylase, which catalyzes the conversion of acetyl-CoA to malonyl-CoA, was replaced with a strong, constitutive promoter (*TEF1* promoter) in a *S. cerevisiae* strain harboring two plasmids carrying the genes encoding 6-MSAS from *P. patulum* and PPTase from *A. nidulans*, respectively. The strain was characterized in batch cultivations with a glucose minimal media, and a 60% increase in 6-MSA titer was observed compared to a strain having the native promoter in front of *ACC1*. However, 6-MSAS gene was also integrated in the chromosome at the site of YMLCΔ2 which resulted in a stable production strain, but yielded less 6-MSA than the plasmid strains.

The strain carrying the 6-MSAS gene with the change of *ACC1* to *TEF1* promoter was cultivated in glucose minimal media with different initial glucose concentration from 20-200 g/L in order to evaluate the optimum glucose concentration for the production of polyketide in yeast. A low production of 6-MSA in both glucose and ethanol phase was observed when an initial glucose concentration between 100-200 g/L was applied, possibly due to the hyperosmotic pressure and toxicity of ethanol as well as other by-products that are formed during the cultivation. The optimum of initial glucose concentration was found between 20-50 g/L as a high specific growth rate and 6-MSA yield were obtained.

Since 6-MSA was used as a model system to develop the platform for the polyketide production in yeast, it is interesting to understand the metabolic regulation of the whole cell during the biosynthesis of 6-MSA at the transcriptional level. Consequently, a DNA expression analysis using Affymetrix gene chips was performed. The evaluation of whole genome transcription response using reporter metabolite and subnetwork analysis revealed the global effect due to the expression of the 6-MSAS gene, not only on the central metabolism, EMP, TCA and PP pathways, but also on the amino acid metabolism when the strain carrying 6-MSA was cultivated on glucose minimal media, and on the fatty acid metabolism when the promoter of *ACC1* was changed to *TEF1* promoter. The result from metabolic flux analysis of [1-^{13}C]-labeled glucose cultivations with the reference strain and strains carrying the 6-MSAS gene was in accordance with the result from transcription analysis and revealed that flux towards acetyl-CoA was increased in the strains carrying the 6-MSAS gene, especially in the strain with the change of the promoter of *ACC1* to *TEF1* promoter.

The identification of further targets for metabolic engineering using a genome scale metabolic model and MOMA approach in order to improve 6-MSA production in yeast revealed that *GDH1* and *SER1* could be the targets for gene deletion, which may yield a 2.3 times higher 6-MSA production than observed for the so far highest yielding strain, having the *ACC1* promoter replaced.

All knowledge generated in this study can definitely be used as a basis for forming a platform for the production of polyketides in the future. For further investigations, the deletion of the target genes as mentioned above must be performed to validate the potential increase in 6-MSA production. The expression of the 6-MSAS gene from *A. terreus* in yeast must be investigated. The sequence of *atX* should be adapted in order to remove the codon bias and unfavorable mRNA secondary structure, and if it is expressed but the production of 6-MSA can not be detected, the experiment to extract 6-MSA out of the cells should be carried out. Furthermore, the test of this yeast system using another fungal polyketide or even more complex polyketides from different organisms is strongly suggested.

Appendix A

Distribution of Transcripts in Gene Ontology (GO) Terms that are Significantly Changed in Different Pair-wise Comparisons

---

Appendix A.1. Distribution of transcripts in Gene Ontology (GO) terms that are significantly changed of the strain IBT100082 in ethanol phase compared to the same strain in glucose phase.

A) Up-regulated genes (499 genes)

GO term	Cluster frequency (total 499 genes)		Background frequency (total 7292 genes)		P-value
	genes	%	genes	%	
generation of precursor metabolites and energy	53	10.6	181	2.5	8.22E-18
energy derivation by oxidation of organic compounds	40	8.0	143	2.0	3.39E-12
oxidative phosphorylation	22	4.4	46	0.6	2.31E-11
coenzyme metabolic process	36	7.2	136	1.9	5.37E-10
cofactor metabolic process	40	8.0	170	2.3	1.71E-09
aerobic respiration	26	5.2	84	1.2	2.07E-08
cellular respiration	26	5.2	90	1.2	1.16E-07
cellular carbohydrate metabolic process	41	8.2	213	2.9	7.52E-07
electron transport	15	3.0	33	0.5	8.19E-07
carbohydrate metabolic process	43	8.6	233	3.2	1.13E-06
ATP synthesis coupled electron transport	13	2.6	25	0.3	1.36E-06
organelle ATP synthesis coupled electron transport	13	2.6	25	0.3	1.36E-06
phosphorylation	33	6.6	155	2.1	2.55E-06
acetyl-CoA metabolic process	11	2.2	20	0.3	1.19E-05
monocarboxylic acid metabolic process	28	5.6	125	1.7	1.29E-05
cofactor catabolic process	11	2.2	21	0.3	2.36E-05
acetyl-CoA catabolic process	9	1.8	15	0.2	9.60E-05
tricarboxylic acid cycle	9	1.8	15	0.2	9.60E-05
cellular catabolic process	58	11.6	421	5.8	0.00013
coenzyme catabolic process	10	2.0	20	0.3	0.00018
phosphorus metabolic process	36	7.2	212	2.9	0.00024
phosphate metabolic process	36	7.2	212	2.9	0.00024
aldehyde metabolic process	10	2.0	21	0.3	0.00033
catabolic process	58	11.6	434	6.0	0.00037
glyoxylate metabolic process	5	1.0	5	0.1	0.00134
TCA intermediate metabolic process	9	1.8	20	0.3	0.00236
carboxylic acid metabolic process	43	8.6	313	4.3	0.00631
organic acid metabolic process	43	8.6	313	4.3	0.00631
mitochondrial electron transport	6	1.2	9	0.1	0.00636

B) Down-regulated genes (436 genes)

GO term	Cluster frequency (total 436 genes)		Background frequency (total 7292 genes)		P-value
	genes	%	genes	%	
ribosome biogenesis and assembly	125	28.7	410	5.6	3.29E-56
ribonucleoprotein complex biogenesis and assembly	129	29.6	483	6.6	9.63E-51
rRNA metabolic process	74	17.0	256	3.5	1.97E-29
rRNA processing	71	16.3	249	3.4	1.10E-27
cellular component organization and biogenesis	249	57.1	2281	31.3	1.12E-27
organelle organization and biogenesis	182	41.7	1395	19.1	1.69E-26
RNA processing	98	22.5	491	6.7	1.35E-25
RNA metabolic process	152	34.9	1056	14.5	1.86E-25
nucleobase and nucleic acid metabolic process	192	44.0	1675	23.0	6.81E-21
cellular process	369	84.6	4670	64.0	6.80E-20
metabolic process	296	67.9	3583	49.1	2.59E-13
primary metabolic process	280	64.2	3317	45.5	3.83E-13
cellular metabolic process	284	65.1	3473	47.6	2.37E-11
ribosomal large subunit biogenesis and assembly	24	5.5	64	0.9	6.11E-11
maturation of SSU-rRNA	20	4.6	45	0.6	1.69E-10
maturation of SSU-rRNA from tricistronic rRNA transcript	20	4.6	45	0.6	1.69E-10
biopolymer metabolic process	208	47.7	2344	32.1	2.15E-09
macromolecule metabolic process	244	56.0	2923	40.1	3.65E-09
maturation of 5.8S rRNA from tricistronic rRNA transcript	15	3.4	34	0.5	2.20E-07
maturation of 5.8S rRNA	15	3.4	36	0.5	5.91E-07
tRNA processing	22	5.0	84	1.2	2.05E-06
tRNA modification	17	3.9	52	0.7	3.54E-06
tRNA metabolic process	26	6.0	121	1.7	6.99E-06
endonucleolytic cleavage in 5'-ETS of tricistronic rRNA transcript	10	2.3	18	0.2	1.44E-05
endonucleolytic cleavage to generate mature 5'-end of SSU-rRNA	10	2.3	19	0.3	2.88E-05
tRNA methylation	9	2.1	15	0.2	3.14E-05
endonucleolytic cleavage in ITS1	10	2.3	20	0.3	5.46E-05
endonucleolytic cleavages during rRNA processing	10	2.3	20	0.3	5.46E-05
endonucleolytic cleavage of tricistronic rRNA transcript	10	2.3	20	0.3	5.46E-05
cleavages during rRNA processing	10	2.3	21	0.3	9.88E-05
ribosome export from nucleus	11	2.5	27	0.4	0.00016
ribosome assembly	17	3.9	66	0.9	0.00019
transcription from RNA polymerase I promoter	12	2.8	34	0.5	0.00028
protein export from nucleus	15	3.4	55	0.8	0.00044
macromolecular complex assembly	43	9.9	329	4.5	0.00072
RNA modification	24	5.5	139	1.9	0.00193
macromolecule localization	46	10.6	384	5.3	0.00345
cellular component assembly	53	12.2	469	6.4	0.00354
nuclear transport	22	5.0	128	1.8	0.00559
nucleocytoplasmic transport	22	5.0	128	1.8	0.00559
nuclear export	19	4.4	101	1.4	0.00629

Appendix A.2. Distribution of transcripts in Gene Ontology (GO) terms that are significantly changed in the strain IBT100082 compared to the reference strain (IBT100081), in the batch cultivation during the ethanol phase.

A) Up-regulated genes (322 genes)

GO term	Cluster frequency (total 322 genes)		Background frequency (total 7292 genes)		P-value
	genes	%	genes	%	
nitrogen compound metabolic process	39	12.1	251	3.4	2.46E-09
amino acid biosynthetic process	24	7.5	106	1.5	1.55E-08
amine biosynthetic process	24	7.5	114	1.6	8.01E-08
nitrogen compound biosynthetic process	24	7.5	115	1.6	9.72E-08
carboxylic acid metabolic process	41	12.7	313	4.3	1.73E-07
organic acid metabolic process	41	12.7	313	4.3	1.73E-07
amine metabolic process	34	10.6	228	3.1	2.14E-07
amino acid metabolic process	30	9.3	187	2.6	4.28E-07
amino acid and derivative metabolic process	30	9.3	202	2.8	2.87E-06
glutamine family amino acid metabolic process	13	4.0	43	0.6	0.0000175
arginine biosynthetic process	7	2.2	10	0.1	0.0000274
glutamine family amino acid biosynthetic process	10	3.1	27	0.4	0.0000884
urea cycle intermediate metabolic process	7	2.2	15	0.2	0.00121
arginine metabolic process	7	2.2	15	0.2	0.00121
cellular carbohydrate metabolic process	26	8.1	213	2.9	0.00171
carbohydrate transport	10	3.1	36	0.5	0.00187
asparagine biosynthetic process	4	1.2	4	0.1	0.00312
heterocycle metabolic process	14	4.3	76	1.0	0.00387
monosaccharide transport	8	2.5	24	0.3	0.00438
hexose transport	8	2.5	24	0.3	0.00438
carbohydrate metabolic process	26	8.1	233	3.2	0.00904
nonprotein amino acid metabolic process	5	1.6	9	0.1	0.01479
nonprotein amino acid biosynthetic process	4	1.2	6	0.1	0.04357
cytokinesis, completion of separation	5	1.6	11	0.2	0.05040

B) Down-regulated genes (395 genes)

GO term	Cluster frequency (total 395 genes)		Background frequency (total 7292 genes)		P-value
	genes	%	genes	%	
DNA repair	27	6.8	192	2.6	0.00382
response to DNA damage stimulus	30	7.6	236	3.2	0.00899
response to endogenous stimulus	30	7.6	245	3.4	0.01889

Appendix A.3. Distribution of transcripts in Gene Ontology (GO) terms that are significantly changed in the strain IBT100082 compared to the reference strain (IBT100081), in the batch cultivation during the glucose phase.

A) Up-regulated genes (600 genes)

GO term	Cluster frequency (total 600 genes)		Background frequency (total 7292 genes)		P-value
	genes	%	genes	%	
carbohydrate metabolic process	51	8.5	233	3.2	4.83E-08
cellular carbohydrate metabolic process	45	7.5	213	2.9	2.39E-06
catabolic process	68	11.3	434	6.0	0.00011
cellular catabolic process	66	11.0	421	5.8	0.00017
carboxylic acid metabolic process	53	8.8	313	4.3	0.00026
organic acid metabolic process	53	8.8	313	4.3	0.00026
cell communication	42	7.0	224	3.1	0.00030
signal transduction	38	6.3	209	2.9	0.00251
monocarboxylic acid metabolic process	27	4.5	125	1.7	0.00276
monosaccharide metabolic process	22	3.7	92	1.3	0.00412
nitrogen compound metabolic process	42	7.0	251	3.4	0.00679
glucose metabolic process	17	2.8	65	0.9	0.01492
alcohol metabolic process	30	5.0	163	2.2	0.02338
carbohydrate biosynthetic process	18	3.0	75	1.0	0.03003
coenzyme metabolic process	26	4.3	136	1.9	0.04227
hexose metabolic process	19	3.2	85	1.2	0.05238
response to stress	65	10.8	488	6.7	0.05939

B) Down-regulated genes (177 genes)

GO term	Cluster frequency (total 177 genes)		Background frequency (total 7292 genes)		P-value
	genes	%	genes	%	
ribosome biogenesis and assembly	39	22.0	410	5.6	3.83E-11
ribonucleoprotein complex biogenesis and assembly	42	23.7	483	6.6	7.70E-11
cellular process	149	84.2	4670	64.0	1.00E-06
RNA processing	34	19.2	491	6.7	1.06E-05
nucleobase and nucleic acid metabolic process	73	41.2	1675	23.0	2.01E-05
metabolic process	121	68.4	3583	49.1	7.99E-05
cellular metabolic process	117	66.1	3473	47.6	0.00025
primary metabolic process	113	63.8	3317	45.5	0.00032
rRNA processing	21	11.9	249	3.4	0.00034
organelle organization and biogenesis	61	34.5	1395	19.1	0.00049
rRNA metabolic process	21	11.9	256	3.5	0.00053
maturation of SSU-rRNA	9	5.1	45	0.6	0.00062
maturation of SSU-rRNA from tricistronic rRNA transcript	9	5.1	45	0.6	0.00062

B) Down-regulated genes (177 genes) (continued)

GO term	Cluster frequency (total 177 genes)		Background frequency (total 7292 genes)		P-value
	genes	%	genes	%	
tRNA modification	9	5.1	52	0.7	0.00223
cellular component organization and biogenesis	83	46.9	2281	31.3	0.00490
RNA metabolic process	47	26.6	1056	14.5	0.00972
macromolecule metabolic process	98	55.4	2923	40.1	0.01433
protein-RNA complex assembly	13	7.3	144	2.0	0.02683
ribosomal subunit assembly	8	4.5	55	0.8	0.02945

Appendix A.4. Distribution of transcripts in Gene Ontology (GO) terms that are significantly changed in the strain IBT100083 compared to the strain IBT100082, in the batch cultivation during the ethanol phase.

A) Up-regulated genes (346 genes)

GO term	Cluster frequency (total 346 genes)		Background frequency (total 7292 genes)		P-value
	genes	%	genes	%	
response to stress	42	12.1	488	6.7	0.09095

B) Down-regulated genes (449 genes)

GO term	Cluster frequency (total 449 genes)		Background frequency (total 7292 genes)		P-value
	genes	%	genes	%	
cellular process	351	78.2	4670	64.0	1.81E-08
metabolic process	267	59.5	3583	49.1	0.00384
nitrogen compound metabolic process	33	7.3.0	251	3.4	0.02558
cellular metabolic process	255	56.8	3473	47.6	0.03781
amine metabolic process	30	6.7	228	3.1	0.05915

Appendix B

The Sequence of *atX* Encoding 6-MSAS from *A. terrues* IBT12713 before and after Engineering

---

Appendix B.1. The sequence *atX* encoding 6-MSAS from *A. terreus* IBT12173 amplified by RT-PCR

```
1       ATGGAGGTACATGGAGATGAAGTGTTGTCAGTCGACTCTGGCGTCTCAACTCCCCCGTCG
         M  E  V  H  G  D  E  V  L  S  V  D  S  G  V  S  T  P  P  S
61      ACAGGAAGTGGATTTCGGAGGCCACTAGAGACCCCCGGAACAGAAATCGGGAATCTCAAT
         T  G  S  G  F  R  R  P  L  E  T  P  G  T  E  I  G  N  L  N
121     CTTAACCCTCAGAATGAGGTTGCCGTTGTTGGAATGGCCTGCCGGCTTGCCGGGGGCAAT
         L  N  P  Q  N  E  V  A  V  V  G  M  A  C  R  L  A  G  G  N
181     AATTCTCCGGAAGAACTGTGGCAGTCCATTCTAAACAGGAAGGATGCCTCTGGCGAGATC
         N  S  P  E  E  L  W  Q  S  I  L  N  R  K  D  A  S  G  E  I
241     CCAAGCATGCGCTGGGAGCCGTATTACCGTCGGGATATTCGCAACCCCAAGATCCTAGAT
         P  S  M  R  W  E  P  Y  Y  R  R  D  I  R  N  P  K  I  L  D
301     CAAACGACAAAGCGCGGCTACTTCTTGGACCACGTCGAGAATTTTGATGCCGCGTTCTTT
         Q  T  T  K  R  G  Y  F  L  D  H  V  E  N  F  D  A  A  F  F
361     GGCGTTTCCCCCAAAGAGGCCGAGCAGATGGACCCCCAGCAGCGGTTGTCACTTGAGGTG
         G  V  S  P  K  E  A  E  Q  M  D  P  Q  Q  R  L  S  L  E  V
421     ACTTGGGAGGCCCTGGAAGACGCAGGAATCCCACCGCAGAGTTTGTCCGGCTCAGAAACA
         T  W  E  A  L  E  D  A  G  I  P  P  Q  S  L  S  G  S  E  T
481     GCCGTGTTTATGGGAGTCAATTCGGATGATTATTCCAAGCTCTTACTGGAAGATATTCCG
         A  V  F  M  G  V  N  S  D  D  Y  S  K  L  L  L  E  D  I  P
541     AACGTGGAGGCCCGGATGGGCATCGGCACTGCGTACTGCGGAGTCCCGAACCGCATCTCC
         N  V  E  A  R  M  G  I  G  T  A  Y  C  G  V  P  N  R  I  S
601     TACCACCTGAACCTCATGGGACCCAGCACTGCCGTTGATGCCGCCTGTGCCTCCTCTCTC
         Y  H  L  N  L  M  G  P  S  T  A  V  D  A  A  C  A  S  S  L
661     GTTGCCATCCATCACGGACGACAAGCCATCCTGCAAGGGGAGAGCGAAGTCGCTATTGTC
         V  A  I  H  H  G  R  Q  A  I  L  Q  G  E  S  E  V  A  I  V
721     GGAGGAGTCAACGCCCTCTGCGGGCCAGGACTGACTCGCGTACTCGACAAGGCAGGAGCG
         G  G  V  N  A  L  C  G  P  G  L  T  R  V  L  D  K  A  G  A
781     ACCTCCACGGAAGGTCGCTGTCTCTCTTTCGACGAAGATGCGAAGGGCTACGGCCGTGGT
         T  S  T  E  G  R  C  L  S  F  D  E  D  A  K  G  Y  G  R  G
841     GAAGGAGCTGCGGTGGTGATCTTGAAACGGCTGTCCACCGCCATCCGGGACGGAGACCAC
         E  G  A  A  V  V  I  L  K  R  L  S  T  A  I  R  D  G  D  H
901     ATTCGCGCCATCATCAAGGGCAGTGCCGTAGCACAGGATGGCAAAACCAACGGCATCATG
         I  R  A  I  I  K  G  S  A  V  A  Q  D  G  K  T  N  G  I  M
961     GCTCCCAACGCCAAGGCACAAGAGCTTGTGGCGTGGAATGCGCTTCGGACAGCCGGAGTC
         A  P  N  A  K  A  Q  E  L  V  A  W  N  A  L  R  T  A  G  V
```

```
1021    GACCCTCTGACGGTTGGGTATGTGGAAGCTCACGCAACGTCAACCCCTCTTGGCGATCCT
         D  P  L  T  V  G  Y  V  E  A  H  A  T  S  T  P  L  G  D  P
1081    ACCGAGGTCAGCGTCGTCTCAGCAGTCTACGGCAAAGGCAGACCGGAAGGGAATCCTTGC
         T  E  V  S  V  V  S  A  V  Y  G  K  G  R  P  E  G  N  P  C
1141    TTCATTGGCTCTGTCAAACCCAACGTGGGCCATTTGGAAGCGGGCGCTGGCGCCGTCGGT
         F  I  G  S  V  K  P  N  V  G  H  L  E  A  G  A  G  A  V  G
1201    TTCATCAAAGCAGTCATGGCAGTTGAAAAGGCCATTTTCCCCCCACAAACCAACCTGAAG
         F  I  K  A  V  M  A  V  E  K  A  I  F  P  P  Q  T  N  L  K
1261    AGACTCAATTCTCGCATTGACTGGGGCCAAGCCGGAGTGAAGGTCGTCCAGGAGACACTG
         R  L  N  S  R  I  D  W  G  Q  A  G  V  K  V  V  Q  E  T  L
1321    GAATGGCCTGGCAATGAGGATGACGTCCGCCGAGCCGGTGTTTGCTCTTACGGATATGGT
         E  W  P  G  N  E  D  D  V  R  R  A  G  V  C  S  Y  G  Y  G
1381    GGTACGGTCTCCCATGCAATCATCGAGGAGTTTGCGCAACAGCTCCAGCGGCCGACTACC
         G  T  V  S  H  A  I  I  E  E  F  A  Q  Q  L  Q  R  P  T  T
1441    AACACAACCGATGAAGACCCTCTGCCTCGGATTCTTCTCCTGTCGGCACCTCAAGAGAGA
         N  T  T  D  E  D  P  L  P  R  I  L  L  L  S  A  P  Q  E  R
1501    CGCCTTGCTTTGCAGGCACGGACACAGGCCTCCTGGATTGCCGCGGAGGGCAGAAATAGA
         R  L  A  L  Q  A  R  T  Q  A  S  W  I  A  A  E  G  R  N  R
1561    ACCCTGGAGTCGATTGCAACCACCTTGAGCACTCGTCGTGGGCACCATGACTACCGGGCT
         T  L  E  S  I  A  T  T  L  S  T  R  R  G  H  H  D  Y  R  A
1621    GCCATCATCGCAGAGAACCATGATGACGCTGTTCAGAAACTGTCTGACATTGTCAATGGT
         A  I  I  A  E  N  H  D  D  A  V  Q  K  L  S  D  I  V  N  G
1681    AAAGCAGCCGAATGGACGACGTCGAGTCGTGTTCTCGATGCCAGTTGCTCCAAGGACGTG
         K  A  A  E  W  T  T  S  S  R  V  L  D  A  S  C  S  K  D  V
1741    GTGTGGGTTTTCTCCGGTCATGGCGCACAATGGACTGCAATGGCTACGGATCTCCTCAAA
         V  W  V  F  S  G  H  G  A  Q  W  T  A  M  A  T  D  L  L  K
1801    GACATTGTGTTCTATCAAACAATCAGCCGTCTGGACCCGATTGTGGAGCGCGAAATGGGC
         D  I  V  F  Y  Q  T  I  S  R  L  D  P  I  V  E  R  E  M  G
1861    TTCTCGGCATTGCATTCCCTTGCAAGTGGCGATTTCGAATCGTCCATCAAGGTGCAAGTG
         F  S  A  L  H  S  L  A  S  G  D  F  E  S  S  I  K  V  Q  V
1921    CTCACCTATCTCGTACAGGTGGGACTGGCTGCCATCTTGCGCTCGAAGGGATTGGAGCCC
         L  T  Y  L  V  Q  V  G  L  A  A  I  L  R  S  K  G  L  E  P
1981    CAGGCTGTCATCGGTCATTCAGTTGGCGAAATTGCCGCCTCAGTCGCGGCTGGCTGTCTG
         Q  A  V  I  G  H  S  V  G  E  I  A  A  S  V  A  A  G  C  L
2041    ACTGCAGAAGAAGGCGCCCTGATTGTCACCCGCAGAGCAAACCTCTATCGGCGTGTGATG
         T  A  E  E  G  A  L  I  V  T  R  R  A  N  L  Y  R  R  V  M
2101    GGCGCGGGCGCAATGGTTCTCGTCAACATTCCATTTGTCGACATGGAGAAAGAGCTTCAA
         G  A  G  A  M  V  L  V  N  I  P  F  V  D  M  E  K  E  L  Q
2161    GGCCGGACGGACCTCGTGGCCGCCATTGGCTCCTCGCCATCTTCATGTGTTGTTTCCGGT
         G  R  T  D  L  V  A  A  I  G  S  S  P  S  S  C  V  V  S  G
2221    GCCACTGAGGCTGTCCTGGCACTCGTGGAAGACCTCAAGTCTCGTGGTGTCAACGCTTTC
         A  T  E  A  V  L  A  L  V  E  D  L  K  S  R  G  V  N  A  F
2281    CGGGTCAAGACGGATATTCCCTTCCACCACCCGATGCTGGATCAACTGTCCGAGCCCTTG
         R  V  K  T  D  I  P  F  H  H  P  M  L  D  Q  L  S  E  P  L
2341    CGAGAGGCCATGGCAGGGTCCCTGTCGCCACGCAAGCCCAGAGTCCGTCTTTACTCGACG
         R  E  A  M  A  G  S  L  S  P  R  K  P  R  V  R  L  Y  S  T
2401    TCGGCAGAAGACCCACGCAGTATGGTTGCTCGGGATATATATTACTGGACCAGCAACATG
         S  A  E  D  P  R  S  M  V  A  R  D  I  Y  Y  W  T  S  N  M
2461    GTCAACCCGGTCCGGTTGACGGCCGCAGTGCAGGCAGCAGTGGACGATGGCCTGCGATTG
         V  N  P  V  R  L  T  A  A  V  Q  A  A  V  D  D  G  L  R  L
2521    TTCCTTGAAGTCTCTTCTCATCCCATTGTGTCCCACTCTGTCCGAGAGACCATGTTGGAC
         F  L  E  V  S  S  H  P  I  V  S  H  S  V  R  E  T  M  L  D
2581    CTGGGTGTGGAGGACTTCACCGTGACCAACACCATGGCTCGCAATAAGCCTGCCGACAAG
         L  G  V  E  D  F  T  V  T  N  T  M  A  R  N  K  P  A  D  K
2641    ACCATTCTGTCCAGCATTGCGCAGCTTCACTGTCGGGGCGCCGTCGTCAATTGGAAGAAG
         T  I  L  S  S  I  A  Q  L  H  C  R  G  A  V  V  N  W  K  K
```

```
2701    CAGCTGCCGGGCCCTTGGGCGCTGGATGTGCCCTTGACGACCTGGGACCACAAGCCCTAC
         Q  L  P  G  P  W  A  L  D  V  P  L  T  T  W  D  H  K  P  Y
2761    TGGCGGCATATTCACACTGGCCCTATCAGTGCCTCGACTTTGCACGATGTGGACAAACAC
         W  R  H  I  H  T  G  P  I  S  A  S  T  L  H  D  V  D  K  H
2821    ACGCTGTTGGGTCAGCGCGTTCCCGTTGCGGGAGAAACAACTATGGTGTTCACCACCCAG
         T  L  L  G  Q  R  V  P  V  A  G  E  T  T  M  V  F  T  T  Q
2881    ATGGATGACCAGACCAAGCCTTTCCCAGGAAGCCATCCACTGCACGGCTCTGAGATTGTT
         M  D  D  Q  T  K  P  F  P  G  S  H  P  L  H  G  S  E  I  V
2941    CCGGCTGCTGCCCTTGTCAACACTTTCCTGCATGCCACCGGGGCTACCACCCTTTCCAAC
         P  A  A  A  L  V  N  T  F  L  H  A  T  G  A  T  T  L  S  N
3001    ATTACCCTTCGCGTGCCAGTGGCCATCAGCCAGCCGCGCGACATCCAGGTGGTGGTGTCA
         I  T  L  R  V  P  V  A  I  S  Q  P  R  D  I  Q  V  V  V  S
3061    CAGAATCAAATCAAGATCTGCTCCCGTCTCACTCAGAAGGCGGGTTCTGGGGCAGACGAA
         Q  N  Q  I  K  I  C  S  R  L  T  Q  K  A  G  S  G  A  D  E
3121    GGTTCCTGGCTGACACACACTACGGGTCAGTGGGAGGCTGGTGGAAGCAAGAACGCCCCG
         G  S  W  L  T  H  T  T  G  Q  W  E  A  G  G  S  K  N  A  P
3181    GCGCAACTCGATATTGCTGCTATCAAGGCTCGTCTCGCTAATAACAAATTGGCGGACAAC
         A  Q  L  D  I  A  A  I  K  A  R  L  A  N  N  K  L  A  D  N
3241    TTCTCCATCGACTATTTGGACAAGGTTGGCGTTTCAGCAATGGGCTTTCCTTGGGCAGTT
         F  S  I  D  Y  L  D  K  V  G  V  S  A  M  G  F  P  W  A  V
3301    ACAGAGCACTACGGCACCCTGCAGGAGATGATCGCTCGCGTTGATGTCGCGCCAGACGTC
         T  E  H  Y  G  T  L  Q  E  M  I  A  R  V  D  V  A  P  D  V
3361    CCCGCGACCAGTCCACTCCCCTGGGATGCTGCCTCTTGGGCCCCGATCCTCGATGCGGCC
         P  A  T  S  P  L  P  W  D  A  A  S  W  A  P  I  L  D  A  A
3421    ACCTCAGTGGGATCCACTCTCTTTTTCGATCAGCCTCGCCTGCGCATGCCGGCTCACATT
         T  S  V  G  S  T  L  F  F  D  Q  P  R  L  R  M  P  A  H  I
3481    CACGGGGTTCAAGTCTACACCACGCAGCCGCCTCCCAAGGTGGGTTACCTGTACGTGGAA
         H  G  V  Q  V  Y  T  T  Q  P  P  P  K  V  G  Y  L  Y  V  E
3541    AAGGCTGGCGATCGGGATCTGGCGGTGCATGTCAGTGTCTGCGACGAGCTCGGAACCGTC
         K  A  G  D  R  D  L  A  V  H  V  S  V  C  D  E  L  G  T  V
3601    TTAGCTCGATTCGAATCCATGCGCTTTTCCGAGATCGAAGGCACGCCGGGCAGTAACGGC
         L  A  R  F  E  S  M  R  F  S  E  I  E  G  T  P  G  S  N  G
3661    AGCGAGGAGAGTCTTGTCCATCAGCTCGCATGGCCTCCCGCGATCTACAGCGAGAGGCCG
         S  E  E  S  L  V  H  Q  L  A  W  P  P  A  I  Y  S  E  R  P
3721    CTGACAATCAACAATGTCGTCCTCGTTTCCCGGGATAAGAACGTCGCAGATCTCTACTGT
         L  T  I  N  N  V  V  L  V  S  R  D  K  N  V  A  D  L  Y  C
3781    GGGTCCTTGAAAGATCGTGTGTCATCTATCACGGTGCTGGATGCTGCTGCCGACCTGCTT
         G  S  L  K  D  R  V  S  S  I  T  V  L  D  A  A  A  D  L  L
3841    TCCCTTTCGCAGGATTCCAGCAGTGTCTTGCAAGCAAAAGATACAGCGGTGGTGTACGTG
         S  L  S  Q  D  S  S  S  V  L  Q  A  K  D  T  A  V  V  Y  V
3901    CCCGGTCCCCTCCACAGCGCGGATTCTATCCCGACTGCGGCCCATTCTTTCCTCATGGAA
         P  G  P  L  H  S  A  D  S  I  P  T  A  A  H  S  F  L  M  E
3961    TTGCTCCTCCTGGTCAAAATCATTGTCAATGGCTCTTTGCCCACCAAGGTCTTTGTCCTT
         L  L  L  L  V  K  I  I  V  N  G  S  L  P  T  K  V  F  V  L
4021    ACGGACCGCGTCTGCGAGAGTGAGTCTGCTACGGCTCTCGCTCAGTCTCCGATCCACGGT
         T  D  R  V  C  E  S  E  S  A  T  A  L  A  Q  S  P  I  H  G
4081    GTCTCCCGTATCATTGCTGCGGAGCACCCAGATCAATGGGGCGGACTGATTGACGTCGAA
         V  S  R  I  I  A  A  E  H  P  D  Q  W  G  G  L  I  D  V  E
4141    ACGCCGGGCCAGTTCTCACTCGAGACGATGAAGTATGTGCAGGAGGCGGATAACATCCGC
         T  P  G  Q  F  S  L  E  T  M  K  Y  V  Q  E  A  D  N  I  R
4201    ATCTCGGATGGCATACCCAGAATTGCTCGTCTGCGCCCGCTTCCTCGCGACAAGCTCCTA
         I  S  D  G  I  P  R  I  A  R  L  R  P  L  P  R  D  K  L  L
4261    CCGCCTAGCAAGCAGACTTCCCTGCTCCCCCGACCCGAAGGTACCTACTTGATTACGGGT
         P  P  S  K  Q  T  S  L  L  P  R  P  E  G  T  Y  L  I  T  G
4321    GGACTGGGCGCTCTGGGGTTGGAGGTCGCACAGTTCCTGGTGGAAAAGGGTGCTCGTCGA
         G  L  G  A  L  G  L  E  V  A  Q  F  L  V  E  K  G  A  R  R
```

```
4381    TTGATCCTCGTTTCTCGGCGTGCCTTGCCTCCGCGCCGGGAGTGGGCAGACATCCTTGCT
         L  I  L  V  S  R  R  A  L  P  P  R  R  E  W  A  D  I  L  A
4441    GATGCATCATCCTCGCTGGCGCCGGCGCTGGAGACAATCCAGGCCCTTGAAGCACAGGGA
         D  A  S  S  S  L  A  P  A  L  E  T  I  Q  A  L  E  A  Q  G
4501    GCCACTGTCCACACTCTTGCAGTGGACATTTCCTCTCCTGACGCAGCGCCTCAACTGGCT
         A  T  V  H  T  L  A  V  D  I  S  S  P  D  A  A  P  Q  L  A
4561    GTCGCCATTGATTCTCTGTCGCTACCCCCAGTCCGCGGCGTGGTCCACGCAGCAGGCGTT
         V  A  I  D  S  L  S  L  P  P  V  R  G  V  V  H  A  A  G  V
4621    CTGGACAGCCAGCTGGTCCTCTCCGCCACGTCAGACTCTGTCGAGCGCGTGCTGGCGCCC
         L  D  S  Q  L  V  L  S  A  T  S  D  S  V  E  R  V  L  A  P
4681    AAGATCACCGGAGCGCTGGTCCTTGGCACCGTCTTCCCCCCCAAGGCCCTCGATTTCTTC
         K  I  T  G  A  L  V  L  G  T  V  F  P  P  K  A  L  D  F  F
4741    ATGCTATTCTCCTCGTGCGGACAGCTACTAGGCTTCCCGGGTCAAGCATCCTACGCGTCC
         M  L  F  S  S  C  G  Q  L  L  G  F  P  G  Q  A  S  Y  A  S
4801    GGAAACGCGTTCCTTGATGCATTCGCAACCTCGCGCCGACACCAAGGAGACAACGCTGTC
         G  N  A  F  L  D  A  F  A  T  S  R  R  H  Q  G  D  N  A  V
4861    GCCGTGCAGTGGACCAGCTGGCGCTCCCTCGGCATGGCAGCCAGTACCGACTTCATCAAC
         A  V  Q  W  T  S  W  R  S  L  G  M  A  A  S  T  D  F  I  N
4921    GCTGAGCTGGCCAGCAAGGGCATCACTGACATCACGCGCGACGAGGGATTCCGCGCATGG
         A  E  L  A  S  K  G  I  T  D  I  T  R  D  E  G  F  R  A  W
4981    ATGCATATTTCCAAATATGATATCGACCAGGCCGCGGTCTTGCGCAGTCTGGCCTTCGAG
         M  H  I  S  K  Y  D  I  D  Q  A  A  V  L  R  S  L  A  F  E
5041    GCCGATGAACCCCTCCCCACCCCTATCCTTACGGATATTGCCGTCCGCAAGGCTGGCTCC
         A  D  E  P  L  P  T  P  I  L  T  D  I  A  V  R  K  A  G  S
5101    GCCTCCTCCGCGGATGCTCCCTCTGCTGCACCGAAAGAGACGAACGAAATGCCGGAATCG
         A  S  S  A  D  A  P  S  A  A  P  K  E  T  N  E  M  P  E  S
5161    ATCCCGGAGCGTCGTACCTGGTTGGATGAGCGAATCCGTGATTGTGTGGCCCGCGTGCTT
         I  P  E  R  R  T  W  L  D  E  R  I  R  D  C  V  A  R  V  L
5221    CAGCTGGGGAGCAGCGACGAGGTTGATTCCAAGGCCGCTCTGAGTGACCTGGGAGTCGAC
         Q  L  G  S  S  D  E  V  D  S  K  A  A  L  S  D  L  G  V  D
5281    AGTGTCATGACCGTTAGCTTGAGAGGTCAGCTGCAGAAGACGTTGGGGGTCAAGGTGCCA
         S  V  M  T  V  S  L  R  G  Q  L  Q  K  T  L  G  V  K  V  P
5341    CCCACACTGACCTGGAGTTGCCCGACGGTGTCACATCTGGTGGGATGGTTTTTGGAAAAG
         P  T  L  T  W  S  C  P  T  V  S  H  L  V  G  W  F  L  E  K
5401    ATGGGAAATCATCATCATCATCATTAA
         M  G  N  H  H  H  H  H  H  -
```

Appendix B.2. The engineered *atX* encoding 6-MSAS proposed by Java Codon Adaptation Tool

```
1       ATGGAAGTTCACGGTGACGAAGTTTTGTCTGTTGACTCTGGTGTTTCTACTCCACCATCT
         M  E  V  H  G  D  E  V  L  S  V  D  S  G  V  S  T  P  P  S
61      ACTGGTTCTGGTTTCAGAAGACCATTGGAAACTCCAGGTACTGAAATCGGTAACTTGAAC
         T  G  S  G  F  R  R  P  L  E  T  P  G  T  E  I  G  N  L  N
121     TTGAACCCACAAAACGAAGTTGCTGTTGTTGGTATGGCTTGTAGATTGGCTGGTGGTAAC
         L  N  P  Q  N  E  V  A  V  V  G  M  A  C  R  L  A  G  G  N
181     AACTCTCCAGAAGAATTGTGGCAATCTATCTTGAACAGAAAGGACGCTTCTGGTGAAATC
         N  S  P  E  E  L  W  Q  S  I  L  N  R  K  D  A  S  G  E  I
241     CCATCTATGAGATGGGAACCATACTACAGAAGAGACATCAGAAACCCAAAGATCTTGGAC
         P  S  M  R  W  E  P  Y  Y  R  R  D  I  R  N  P  K  I  L  D
301     CAAACTACTAAGAGAGGTTACTTCTTGGACCACGTTGAAAACTTTGATGCTGCGTTCTTC
         Q  T  T  K  R  G  Y  F  L  D  H  V  E  N  F  D  A  A  F  F
361     GGTGTTTCGCCAAAGGAAGCTGAACAAATGGACCCACAACAAAGATTGTCTTTGGAAGTT
         G  V  S  P  K  E  A  E  Q  M  D  P  Q  Q  R  L  S  L  E  V
421     ACTTGGGAAGCTTTGGAAGACGCTGGTATCCCACCACAATCTTTGTCTGGTTCTGAAACT
         T  W  E  A  L  E  D  A  G  I  P  P  Q  S  L  S  G  S  E  T
481     GCTGTTTTCATGGGTGTTAACTCTGACGACTACTCTAAGTTGTTGTTGGAAGACATCCCA
         A  V  F  M  G  V  N  S  D  D  Y  S  K  L  L  L  E  D  I  P
541     AACGTTGAAGCTAGAATGGGTATCGGTACTGCTTACTGTGGTGTTCCAAACAGAATCTCT
         N  V  E  A  R  M  G  I  G  T  A  Y  C  G  V  P  N  R  I  S
601     TACCACTTGAACTTGATGGGTCCATCTACTGCTGTTGACGCTGCTTGTGCTTCTTCTTTG
         Y  H  L  N  L  M  G  P  S  T  A  V  D  A  A  C  A  S  S  L
661     GTTGCTATCCACCACGGTAGACAAGCTATCTTGCAAGGTGAATCTGAAGTTGCTATCGTT
         V  A  I  H  H  G  R  Q  A  I  L  Q  G  E  S  E  V  A  I  V
721     GGTGGTGTTAACGCTTTGTGTGGTCCAGGTTTGACTAGAGTTTTGGACAAGGCTGGTGCT
         G  G  V  N  A  L  C  G  P  G  L  T  R  V  L  D  K  A  G  A
781     ACTTCTACTGAAGGCCGTTGCCTATCTTTCGACGAAGACGCTAAGGGTTACGGTAGAGGT
         T  S  T  E  G  R  C  L  S  F  D  E  D  A  K  G  Y  G  R  G
841     GAAGGTGCTGCTGTTGTTATCTTGAAGAGATTGTCTACTGCTATCAGAGACGGTGACCAC
         E  G  A  A  V  V  I  L  K  R  L  S  T  A  I  R  D  G  D  H
901     ATCAGAGCTATCATCAAGGGTTCTGCTGTTGCTCAAGACGGTAAGACTAACGGTATCATG
         I  R  A  I  I  K  G  S  A  V  A  Q  D  G  K  T  N  G  I  M
961     GCTCCAAACGCTAAGGCTCAAGAATTGGTTGCTTGGAACGCTTTGAGAACTGCTGGTGTT
         A  P  N  A  K  A  Q  E  L  V  A  W  N  A  L  R  T  A  G  V
1021    GACCCATTGACTGTTGGTTACGTTGAAGCTCACGCTACTTCTACTCCATTGGGTGACCCA
         D  P  L  T  V  G  Y  V  E  A  H  A  T  S  T  P  L  G  D  P
1081    ACTGAAGTGAGCGTTGTTTCGGCTGTGTACGGAAAGGGTAGACCAGAAGGTAACCCATGT
         T  E  V  S  V  V  S  A  V  Y  G  K  G  R  P  E  G  N  P  C
1141    TTCATCGGTTCTGTTAAGCCAAACGTTGGTCACTTGGAAGCTGGTGCTGGTGCTGTTGGT
         F  I  G  S  V  K  P  N  V  G  H  L  E  A  G  A  G  A  V  G
1201    TTCATCAAGGCTGTTATGGCTGTTGAAAAGGCTATCTTCCCACCACAAACTAACTTGAAG
         F  I  K  A  V  M  A  V  E  K  A  I  F  P  P  Q  T  N  L  K
1261    AGATTGAACTCTAGAATCGACTGGGGTCAAGCTGGTGTTAAGGTTGTTCAAGAAACTTTG
         R  L  N  S  R  I  D  W  G  Q  A  G  V  K  V  V  Q  E  T  L
1321    GAATGGCCAGGTAACGAAGACGACGTTAGAAGAGCTGGTGTTTGTTCTTACGGTTACGGC
         E  W  P  G  N  E  D  D  V  R  R  A  G  V  C  S  Y  G  Y  G
1381    GGCACTGTAAGCCACGCTATCATCGAAGAATTCGCTCAACAATTGCAAAGACCAACTACT
         G  T  V  S  H  A  I  I  E  E  F  A  Q  Q  L  Q  R  P  T  T
1441    AACACTACTGACGAAGACCCATTGCCAAGAATCTTGTTGTTGTCTGCTCCACAAGAAAGA
         N  T  T  D  E  D  P  L  P  R  I  L  L  L  S  A  P  Q  E  R
1501    AGATTGGCTTTGCAAGCTAGAACTCAAGCTTCTTGGATCGCTGCTGAAGGTAGAAACAGA
         R  L  A  L  Q  A  R  T  Q  A  S  W  I  A  A  E  G  R  N  R
```

```
1561    ACTTTGGAATCTATCGCTACTACTTTGTCTACTAGAAGAGGTCACCACGACTACAGAGCT
         T  L  E  S  I  A  T  T  L  S  T  R  R  G  H  H  D  Y  R  A
1621    GCTATCATCGCTGAAAACCACGACGACGCTGTTCAAAAGTTGTCTGACATCGTTAACGGT
         A  I  I  A  E  N  H  D  D  A  V  Q  K  L  S  D  I  V  N  G
1681    AAGGCTGCTGAATGGACTACTTCTTCTAGAGTTTTGGACGCTTCTTGTAGTAAGGACGTT
         K  A  A  E  W  T  T  S  S  R  V  L  D  A  S  C  S  K  D  V
1741    GTGTGGGTGTTCTCTGGTCACGGTGCTCAATGGACTGCTATGGCTACTGACTTGTTGAAG
         V  W  V  F  S  G  H  G  A  Q  W  T  A  M  A  T  D  L  L  K
1801    GACATCGTTTTCTACCAAACTATCTCTAGATTGGACCCAATCGTTGAAAGAGAAATGGGT
         D  I  V  F  Y  Q  T  I  S  R  L  D  P  I  V  E  R  E  M  G
1861    TTTAGTGCTCTCCACTCTTTGGCGTCTGGTGACTTCGAATCTTCTATCAAAGTTCAAGTA
         F  S  A  L  H  S  L  A  S  G  D  F  E  S  S  I  K  V  Q  V
1921    CTCACTTACCTCGTTCAGGTTGGTCTCGCTGCTATCCTCAGATCTAAGGGTTTGGAACCA
         L  T  Y  L  V  Q  V  G  L  A  A  I  L  R  S  K  G  L  E  P
1981    CAAGCTGTTATCGGTCACTCTGTTGGTGAAATCGCTGCTTCTGTTGCTGCTGGTTGTTTG
         Q  A  V  I  G  H  S  V  G  E  I  A  A  S  V  A  A  G  C  L
2041    ACTGCTGAAGAAGGTGCTTTGATCGTTACTAGAAGAGCTAACTTGTACAGAAGAGTTATG
         T  A  E  E  G  A  L  I  V  T  R  R  A  N  L  Y  R  R  V  M
2101    GGTGCTGGTGCTATGGTTTTGGTTAACATCCCATTCGTTGACATGGAAAAGGAATTGCAA
        G  A  G  A  M  V  L  V  N  I  P  F  V  D  M  E  K  E  L  Q
2161    GGTAGAACTGACTTGGTTGCTGCTATCGGTTCTTCTCCATCTTCTTGTGTTGTTTCTGGT
         G  R  T  D  L  V  A  A  I  G  S  S  P  S  S  C  V  V  S  G
2221    GCTACTGAAGCTGTTTTGGCTTTGGTTGAAGACTTGAAGTCTAGAGGTGTTAACGCTTTC
         A  T  E  A  V  L  A  L  V  E  D  L  K  S  R  G  V  N  A  F
2281    AGAGTTAAGACTGACATCCCATTCCACCACCCAATGTTGGACCAATTGTCTGAACCATTG
         R  V  K  T  D  I  P  F  H  H  P  M  L  D  Q  L  S  E  P  L
2341    AGAGAAGCTATGGCTGGTTCTTTGTCTCCAAGAAAGCCAAGAGTTAGATTGTACTCTACT
         R  E  A  M  A  G  S  L  S  P  R  K  P  R  V  R  L  Y  S  T
2401    TCTGCTGAAGACCCAAGATCTATGGTTGCTAGAGACATCTACTACTGGACTTCTAACATG
         S  A  E  D  P  R  S  M  V  A  R  D  I  Y  Y  W  T  S  N  M
2461    GTTAACCCAGTTAGATTGACTGCTGCTGTTCAAGCTGCTGTTGACGACGGTTTGAGATTG
         V  N  P  V  R  L  T  A  A  V  Q  A  A  V  D  D  G  L  R  L
2521    TTCTTGGAAGTTTCTTCTCACCCAATCGTTTCTCACTCTGTTAGAGAAACTATGTTGGAC
         F  L  E  V  S  S  H  P  I  V  S  H  S  V  R  E  T  M  L  D
2581    TTGGGTGTTGAAGACTTCACTGTTACTAACACTATGGCTAGAAACAAGCCAGCTGACAAG
         L  G  V  E  D  F  T  V  T  N  T  M  A  R  N  K  P  A  D  K
2641    ACTATCTTGTCTTCTATCGCTCAATTGCACTGTAGAGGTGCTGTTGTTAACTGGAAGAAG
         T  I  L  S  S  I  A  Q  L  H  C  R  G  A  V  V  N  W  K  K
2701    CAATTGCCAGGTCCATGGGCTTTGGACGTTCCATTGACTACTTGGGACCACAAGCCATAC
         Q  L  P  G  P  W  A  L  D  V  P  L  T  T  W  D  H  K  P  Y
2761    TGGAGACACATCCACACTGGTCCAATCTCTGCTTCTACTTTGCACGACGTTGACAAGCAC
         W  R  H  I  H  T  G  P  I  S  A  S  T  L  H  D  V  D  K  H
2821    ACTTTGTTGGGTCAAAGAGTTCCAGTTGCTGGTGAAACTACTATGGTTTTCACTACTCAA
         T  L  L  G  Q  R  V  P  V  A  G  E  T  T  M  V  F  T  T  Q
2881    ATGGACGACCAAACTAAGCCATTCCCAGGTTCTCACCCATTGCACGGTTCTGAAATCGTT
         M  D  D  Q  T  K  P  F  P  G  S  H  P  L  H  G  S  E  I  V
2941    CCAGCTGCTGCTTTGGTTAACACTTTCTTGCACGCTACTGGTGCTACTACTTTGTCTAAC
         P  A  A  A  L  V  N  T  F  L  H  A  T  G  A  T  T  L  S  N
3001    ATCACTTTGAGAGTTCCAGTTGCTATCTCTCAACCAAGAGACATCCAAGTTGTTGTTTCT
         I  T  L  R  V  P  V  A  I  S  Q  P  R  D  I  Q  V  V  V  S
3061    CAAAACCAAATCAAGATCTGTTCTAGATTGACTCAAAAGGCTGGTTCTGGTGCTGACGAA
         Q  N  Q  I  K  I  C  S  R  L  T  Q  K  A  G  S  G  A  D  E
3121    GGTTCTTGGTTGACTCACACTACTGGTCAATGGGAAGCTGGTGGTTCTAAGAACGCTCCA
         G  S  W  L  T  H  T  T  G  Q  W  E  A  G  G  S  K  N  A  P
3181    GCTCAATTGGACATCGCTGCTATCAAGGCTAGATTGGCTAACAACAAGTTGGCTGACAAC
         A  Q  L  D  I  A  A  I  K  A  R  L  A  N  N  K  L  A  D  N
```

```
3241    TTCTCTATCGACTACTTGGACAAGGTTGGTGTTTCTGCTATGGGTTTCCCATGGGCTGTT
         F  S  I  D  Y  L  D  K  V  G  V  S  A  M  G  F  P  W  A  V
3301    ACTGAACACTACGGTACTTTGCAAGAAATGATCGCTAGAGTTGACGTTGCTCCAGACGTT
         T  E  H  Y  G  T  L  Q  E  M  I  A  R  V  D  V  A  P  D  V
3361    CCAGCTACTTCTCCATTGCCATGGGACGCTGCTTCTTGGGCTCCAATCTTGGACGCTGCT
         P  A  T  S  P  L  P  W  D  A  A  S  W  A  P  I  L  D  A  A
3421    ACTTCTGTTGGTTCTACTTTGTTCTTCGACCAACCAAGATTGAGAATGCCAGCTCACATC
         T  S  V  G  S  T  L  F  F  D  Q  P  R  L  R  M  P  A  H  I
3481    CACGGTGTTCAAGTTTACACTACTCAACCACCACCAAAGGTTGGTTACTTGTACGTTGAA
         H  G  V  Q  V  Y  T  T  Q  P  P  P  K  V  G  Y  L  Y  V  E
3541    AAGGCTGGTGACAGAGACCTCGCGGTTCACGTAAGCGTGTGCGACGAATTGGGTACTGTT
         K  A  G  D  R  D  L  A  V  H  V  S  V  C  D  E  L  G  T  V
3601    TTGGCTAGATTCGAATCTATGAGATTCTCTGAAATCGAAGGTACTCCAGGTTCTAACGGT
         L  A  R  F  E  S  M  R  F  S  E  I  E  G  T  P  G  S  N  G
3661    TCTGAGGAGAGCTTGGTTCACCAACTCGCTTGGCCACCAGCTATCTACTCTGAAAGACCA
         S  E  E  S  L  V  H  Q  L  A  W  P  P  A  I  Y  S  E  R  P
3721    TTGACTATCAATAATGTTGTGCTGGTTTCTAGAGACAAGAACGTTGCTGACTTGTACTGT
         L  T  I  N  N  V  V  L  V  S  R  D  K  N  V  A  D  L  Y  C
3781    GGTTCGTTGAAGGACAGAGTGTCGTCGATCACTGTGCTAGACGCTGCTGCTGACTTGTTG
         G  S  L  K  D  R  V  S  S  I  T  V  L  D  A  A  A  D  L  L
3841    TCTTTGTCTCAAGACTCTTCTTCTGTTTTGCAAGCTAAGGACACTGCTGTTGTTTACGTT
         S  L  S  Q  D  S  S  S  V  L  Q  A  K  D  T  A  V  V  Y  V
3901    CCAGGTCCATTGCACTCTGCTGACTCTATCCCAACTGCTGCTCACTCTTTCTTGATGGAA
         P  G  P  L  H  S  A  D  S  I  P  T  A  A  H  S  F  L  M  E
3961    TTGTTGTTGTTGGTTAAGATCATCGTTAACGGTTCTTTGCCAACTAAGGTTTTCGTTTTG
         L  L  L  L  V  K  I  I  V  N  G  S  L  P  T  K  V  F  V  L
4021    ACTGACAGAGTTTGTGAATCTGAATCTGCTACTGCTTTGGCTCAATCTCCAATCCACGGT
         T  D  R  V  C  E  S  E  S  A  T  A  L  A  Q  S  P  I  H  G
4081    GTTTCTAGAATCATCGCTGCTGAACACCCCGACCAATGGGGTGGTCTCATCGACGTTGAG
         V  S  R  I  I  A  A  E  H  P  D  Q  W  G  G  L  I  D  V  E
4141    ACTCCAGGTCAATTCTCTTTGGAAACTATGAAGTACGTTCAAGAAGCTGACAACATCAGA
         T  P  G  Q  F  S  L  E  T  M  K  Y  V  Q  E  A  D  N  I  R
4201    ATCTCTGACGGTATCCCAAGAATCGCTAGATTGAGACCATTGCCAAGAGACAAGTTGTTG
         I  S  D  G  I  P  R  I  A  R  L  R  P  L  P  R  D  K  L  L
4261    CCACCATCTAAGCAAACTTCTTTGTTGCCAAGACCAGAAGGTACTTACTTGATCACTGGT
         P  P  S  K  Q  T  S  L  L  P  R  P  E  G  T  Y  L  I  T  G
4321    GGTTTGGGTGCTTTGGGTTTGGAAGTTGCTCAATTCTTGGTTGAAAAGGGTGCTAGAAGA
         G  L  G  A  L  G  L  E  V  A  Q  F  L  V  E  K  G  A  R  R
4381    TTGATCTTGGTTTCTAGAAGAGCTTTGCCACCAAGAAGAGAATGGGCTGACATCTTGGCT
         L  I  L  V  S  R  R  A  L  P  P  R  R  E  W  A  D  I  L  A
4441    GACGCTTCTTCTTCTTTGGCTCCAGCTTTGGAAACTATCCAAGCTTTGGAAGCTCAAGGT
         D  A  S  S  S  L  A  P  A  L  E  T  I  Q  A  L  E  A  Q  G
4501    GCTACTGTTCACACTTTGGCTGTTGACATCTCTTCTCCAGACGCTGCTCCACAATTGGCT
         A  T  V  H  T  L  A  V  D  I  S  S  P  D  A  A  P  Q  L  A
4561    GTTGCTATCGACTCTTTGTCTTTGCCACCAGTTAGAGGTGTTGTTCACGCTGCTGGTGTT
         V  A  I  D  S  L  S  L  P  P  V  R  G  V  V  H  A  A  G  V
4621    TTGGACTCTCAATTGGTTTTGTCTGCTACTTCTGACTCGGTTGAGAGAGTCCTGGCTCCA
         L  D  S  Q  L  V  L  S  A  T  S  D  S  V  E  R  V  L  A  P
4681    AAGATCACTGGCGCGTTGGTACTAGGCACTGTATTCCCACCGAAGGCTTTGGACTTCTTC
         K  I  T  G  A  L  V  L  G  T  V  F  P  P  K  A  L  D  F  F
4741    ATGTTGTTCTCTTCTTGTGGTCAATTGTTGGGTTTCCCAGGTCAAGCTTCTTACGCTTCT
         M  L  F  S  S  C  G  Q  L  L  G  F  P  G  Q  A  S  Y  A  S
4801    GGTAACGCTTTCTTGGACGCTTTCGCTACTTCTAGAAGACACCAAGGTGACAACGCTGTT
         G  N  A  F  L  D  A  F  A  T  S  R  R  H  Q  G  D  N  A  V
4861    GCTGTTCAATGGACTTCTTGGAGATCTTTGGGTATGGCTGCTTCTACTGACTTCATCAAC
         A  V  Q  W  T  S  W  R  S  L  G  M  A  A  S  T  D  F  I  N
```

```
4921    GCTGAATTGGCTTCTAAGGGTATCACTGACATCACTAGAGACGAAGGTTTCAGAGCTTGG
         A  E  L  A  S  K  G  I  T  D  I  T  R  D  E  G  F  R  A  W
4981    ATGCACATCTCTAAGTACGACATCGACCAAGCTGCTGTTTTGAGATCTTTGGCTTTCGAA
         M  H  I  S  K  Y  D  I  D  Q  A  A  V  L  R  S  L  A  F  E
5041    GCTGACGAACCATTGCCAACTCCAATCTTGACTGACATCGCTGTTAGAAAGGCTGGTTCT
         A  D  E  P  L  P  T  P  I  L  T  D  I  A  V  R  K  A  G  S
5101    GCTTCTTCTGCTGACGCTCCATCTGCTGCTCCAAAGGAAACTAACGAAATGCCAGAATCT
         A  S  S  A  D  A  P  S  A  A  P  K  E  T  N  E  M  P  E  S
5161    ATCCCAGAAAGAAGAACTTGGTTGGACGAAAGAATCAGAGACTGTGTTGCTAGAGTTTTG
         I  P  E  R  R  T  W  L  D  E  R  I  R  D  C  V  A  R  V  L
5221    CAATTGGGTTCTTCTGACGAAGTTGACTCTAAGGCTGCTTTGTCTGACTTGGGTGTTGAC
         Q  L  G  S  S  D  E  V  D  S  K  A  A  L  S  D  L  G  V  D
5281    TCTGTTATGACTGTTTCTTTGAGAGGTCAATTGCAAAAGACTTTGGGTGTTAAGGTTCCA
         S  V  M  T  V  S  L  R  G  Q  L  Q  K  T  L  G  V  K  V  P
5341    CCAACTTTGACTTGGTCTTGTCCAACTGTTTCTCACTTGGTTGGTTGGTTCTTGGAAAAG
         P  T  L  T  W  S  C  P  T  V  S  H  L  V  G  W  F  L  E  K
5401    ATGGGTAACCATCATCATCATCATTAA
         M  G  N  H  H  H  H  H  H  -
```

Appendix C

Codon Usage in *Saccharomyces cerevisiae*

Saccharomyces cerevisiae [gbpln]: 14374 CDS's (6511964 codons)

fields: [triplet] [frequency: per thousand] ([number]) ([relative adaptiveness])

```
UUU 26.1(170060)(100)  UCU 23.4(152620)(100)  UAU 18.8(122354)(100)  UGU  8.1( 52733)(100)
UUC 18.4(119910)( 70)  UCC 14.2( 92548)( 61)  UAC 14.8( 96248)( 79)  UGC  4.8( 30943)( 59)
UUA 26.2(170327)( 96)  UCA 18.7(121596)( 80)  UAA  1.0(  6799)(100)  UGA  0.7(  4295)( 70)
UUG 27.2(176823)(100)  UCG  8.6( 55800)( 37)  UAG  0.5(  3265)( 50)  UGG 10.4( 67560)(100)

CUU 12.3( 79881)( 45)  CCU 13.5( 87990)( 74)  CAU 13.6( 88855)(100)  CGU  6.4( 41669)( 30)
CUC  5.4( 35412)( 20)  CCC  6.8( 44190)( 37)  CAC  7.8( 50602)( 57)  CGC  2.6( 16894)( 12)
CUA 13.4( 87337)( 49)  CCA 18.3(118921)(100)  CAA 27.3(177781)(100)  CGA  3.0( 19515)( 14)
CUG 10.5( 68195)( 39)  CCG  5.3( 34483)( 29)  CAG 12.1( 78912)( 44)  CGG  1.7( 11328)(  8)

AUU 30.1(196148)(100)  ACU 20.2(131748)(100)  AAU 35.7(232548)(100)  AGU 14.2( 92246)( 61)
AUC 17.1(111678)( 57)  ACC 12.7( 82630)( 63)  AAC 24.8(161656)( 69)  AGC  9.7( 63472)( 41)
AUA 17.8(116024)( 59)  ACA 17.8(115724)( 88)  AAA 41.9(272927)(100)  AGA 21.3(138740)(100)
AUG 21.0(136644)(100)  ACG  8.0( 51972)( 40)  AAG 30.8(200886)( 74)  AGG  9.2( 60209)( 43)

GUU 22.1(143671)(100)  GCU 21.2(137737)(100)  GAU 37.6(245031)(100)  GGU 23.9(155345)(100)
GUC 11.7( 76454)( 53)  GCC 12.6( 81955)( 59)  GAC 20.2(131713)( 54)  GGC  9.8( 63684)( 41)
GUA 11.8( 76661)( 53)  GCA 16.2(105545)( 76)  GAA 45.6(297191)(100)  GGA 10.9( 71034)( 46)
GUG 10.8( 70084)( 49)  GCG  6.2( 40267)( 29)  GAG 19.2(125264)( 42)  GGG  6.0( 39230)( 25)
```

Coding GC 39.76% 1st letter GC 44.58% 2nd letter GC 36.62% 3rd letter GC 38.10%

Source: http://www.kazusa.or.jp/codon/